

A:Accession: 147438  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-73 <JOR>  
A:Cross-references: EMBL:AL138649  
A:Experimental source: cultivar C  
C:Genetics:  
A:Map position: 3

A: Note: T14D3.120

Query Match 77.7%; Score 306; DB 2; Length 73;  
Best Local Similarity 79.2%; Pred. No. 1.2e-27;  
Matches 57; Conservative 6; Mismatches 9; Indels 0; Gaps 0;

Qy 1 MIEVVCNDRLGKVRVCKNTDDTIGDLKLLIAAQTGRNKKIVLKKWYTFKDHVSLG DY 60  
Db 1 MIEVVCNDRLGKVRVCKNTDDTIGDLKLLIAAQTGRNKKIVLKKWYTFKDHVSLG DY 60  
Qy 61 EHDGMNLELYY 72  
Db 61 EHDGMNLELYY 72

## RESULT 3

ubiquitin-like protein - fission yeast (Schizosaccharomyces pombe)  
C: Species: Schizosaccharomyces pombe  
C: Date: 03-Dec-1999 #sequence\_revision 03-Dec-1999 #text\_change 03-Dec-1999  
C: Accession: T40200  
R: Callis, K.; Harris, D.; Wood, V.; Rajandream, M.A.; Barrell, B.G.  
A: Title: Structure and evolution of genes encoding polyubiquitin and ubiquitin-like p  
A: Reference number: 221912  
A: Accession: T40200  
A: Status: preliminary; translated from GB/EMBL/DBJ  
A: Molecule type: DNA  
A: Residues: 1-73 <OLI>  
A: Cross-references: EMBL:AL049190; PIDN: CAB39137.1; GSPDB: GN00067; SPDB: SPBC31E1.03  
A: Experimental source: strain 972h; cosmid c31E1  
C: Genetics:  
A: Gene: SPDB: SPBC31E1.03  
A: Map position: 2  
A: Introns: 6/2; 40/2

Query Match 76.4%; Score 301; DB 2; Length 73;  
Best Local Similarity 73.6%; Pred. No. 4.3e-27;  
Matches 53; Conservative 12; Mismatches 7; Indels 0; Gaps 0;

Qy 1 MIEVVCNDRLGKVRVCKNTDDTIGDLKLLIAAQTGRNKKIVLKKWYTFKDHVSLG DY 60  
Db 1 MIEVVCNDRLGKVRVCKMPDVTGDFKLLVAAQTGRPRIVLKKWHSVFKNITLADY 60  
Qy 61 EHDGMNLELYY 72  
Db 61 EHDGMNLELYY 72

## RESULT 4

Protein YNR032c-a - yeast (Saccharomyces cerevisiae)  
C: Species: Saccharomyces cerevisiae  
C: Date: 15-Jan-1999 #sequence\_revision 15-Jan-1999 #text\_change 15-Jan-1999  
C: Accession: S78735  
R: Pohl, T.M.  
A: Title: Molecular phylogeny of the env gene of HIV-1 strains isolated in Cot  
submitted to the Protein Sequence Database, April 1996  
A: Reference number: S63346  
A: Accession: S78735  
A: Status: preliminary  
A: Molecule type: DNA  
A: Residues: 1-73 <POH>  
A: Cross-references: EMBL: Z71647; MIPS: YNR032c-a  
A: Experimental source: strain S288C  
C: Genetics:  
A: Map position: 14R

Query Match 64.0%; Score 252; DB 2; Length 73;  
Best Local Similarity 65.3%; Pred. No. 1.4e-21;  
Matches 47; Conservative 12; Mismatches 13; Indels 0; Gaps 0;

Qy 1 MIEVVCNDRLGKVRVCKNTDDTIGDLKLLIAAQTGRNKKIVLKKWYTFKDHVSLG DY 60  
Db 1 MIEVVCNDRLGKVRVCKLAEDSVGDFKVLQLGIGTQPNKIVLQKGGSVLKDHSLEDY 60  
Qy 61 EHDGMNLELYY 72  
Db 61 EHDGMNLELYY 72

## RESULT 5

S55243  
ubiquitin-like protein 8 - Arabidopsis thaliana  
C: Species: Arabidopsis thaliana (mouse-ear cress)  
C: Date: 10-Oct-1995 #sequence\_revision 03-Nov-1995 #text\_change 24-Sep-1999  
C: Accession: S55243; S61068  
R: Callis, J.; Carpenter, T.; Sun, C.W.; Vierstra, R.D.  
A: Title: Structure and evolution of genes encoding polyubiquitin and ubiquitin-like p  
A: Reference number: S55242; MUID: 95229071  
A: Accession: S55243  
A: Status: nucleic acid sequence not shown  
A: Molecule type: DNA  
A: Residues: 1-631 <CAL>  
A: Cross-references: EMBL: L05917  
A: Experimental source: ecotype Columbia  
R: Callis, J.; Carpenter, T.; Sun, C.W.; Vierstra, R.D.  
A: Title: Structure and evolution of genes encoding polyubiquitin and ubiquitin-like p  
submitted to the EMBL Data Library, June 1995  
A: Reference number: S61068  
A: Accession: S61068  
A: Molecule type: DNA  
A: Residues: 1-341, E', 343-631 <CAW>  
A: Cross-references: EMBL: L05917; NID: g870793; PIDN: AAA68879.1; PID: g870794  
C: Superfamily: unassigned ubiquitin-related proteins; ubiquitin homology  
F: 3-78/Domain: ubiquitin homology <UBH1>  
F: 79-154/Domain: ubiquitin homology <UBH2>  
F: 155-237/Domain: ubiquitin homology <UBH3>  
F: 238-318/Domain: ubiquitin homology <UBH4>  
F: 319-392/Domain: ubiquitin homology <UBH5>  
F: 393-468/Domain: ubiquitin homology <UBH6>  
F: 469-551/Domain: ubiquitin homology <UBH7>  
F: 552-627/Domain: ubiquitin homology <UBH8>

Query Match 20.3%; Score 80; DB 2; Length 631;  
Best Local Similarity 31.7%; Pred. No. 0.32;  
Matches 19; Conservative 16; Mismatches 25; Indels 0; Gaps 0;

Qy 11 GKVRVCKNTDDTIGDLKLLIAAQTGRNKKIVLKKWYTFKDHVSLG DY EHDGMNLEL 70  
Db 561 GKTIIEVSSDTIANVKEIKVQKIPDQQLIFFGQOQLEDGVTGLGDIHKKSTLYL 620

## RESULT 6

S60529  
envelope polyprotein gp41 - human immunodeficiency virus type 1 (isolate CI-14-13) (f  
C: Species: human immunodeficiency virus type 1, HIV-1  
A: Variety: isolate CI-14-13  
C: Date: 20-Jul-1996 #sequence\_revision 13-Mar-1997 #text\_change 26-Aug-1999  
C: Accession: S60529  
R: Janssens, W.; Heyndrickx, L.; Van de Peer, Y.; Bouckaert, A.; Franssen, K.; Motte, J  
AIDS 8, 21-26, 1994  
A: Title: Molecular phylogeny of part of the env gene of HIV-1 strains isolated in Cot  
A: Reference number: S60521; MUID: 94280700  
A: Accession: S60529  
A: Status: nucleic acid sequence not shown; translation not shown  
A: Molecule type: DNA  
A: Residues: 1-299 <JAN>  
A: Cross-references: EMBL: X72031; NID: g468637; PIDN: CAA50914.1; PID: g468638  
A: Experimental source: isolate CI-14-13  
A: Note: the nucleotide sequence was submitted to the EMBL Data Library, May 1993  
C: Genetics:  
A: Gene: env  
C: Superfamily: type E retrovirus env polyprotein

RESULT 8  
T29404  
hypothetical protein C16C8.4 - *Caenorhabditis elegans*  
C:Species: *Caenorhabditis elegans*  
C>Date: 15-Oct-1999 #sequence\_revision 15-Oct-1999 #text\_change 15-Oct-1999  
C:Accession: T29404  
R:Waterston, R.; Le, T. T.; Gattung, S.  
Submitted to the EMBL Data Library, November 1996  
A:Description: The sequence of *C. elegans* cosmid C16C8.  
A:Reference number: Z20617  
A:Accession: T29404  
A:Status: preliminary; translated from GB/EMBL/DBJ  
A:Molecule type: DNA

S60345  
envelope glycoprotein gp41 - human immunodeficiency virus type 1 (isolate CI-45-1) (fr  
C;Species: human immunodeficiency virus type 1, HIV-1  
A;Variety: isolate CI-45-1  
C;Date: 20-Jul-1996 #sequence\_revision 13-Mar-1997 #text\_change 26-Aug-1999  
C;Accession: S60345  
R;Janssens, W.; Heyndrickx, L.; Van de Peer, Y.; Bouckaert, A.; Fransen, K.; Motte, J.  
AIDS 8, 21-26, 1994  
A;Title: Molecular phylogeny of part of the env gene of HIV-1 strains isolated in Co  
A;Reference number: S60321; MUID:94280700  
A;Accession: S60345  
A;Status: nucleic acid sequence not shown; translation not shown  
A;Molecule type: DNA  
A;Residues: 1-294 <JAN>  
A;Cross-references: EMBL:X72047; NID:g468669; PIDN:CAA50930.1; PTD:g468670  
A;Experimental source: isolate CI-45-1

A:Note: the nucleotide sequence was submitted to the EMBL Data Library, May 1993

C:Genetics:

A:Gene: env

C:Superfamily: type E retrovirus env polyprotein

C:Keywords: AIDS; glycoprotein; immunodeficiency; polyprotein

Query Match 17.3%; Score 68; DB 2; Length 294;  
Best Local Similarity 26.8%; Pred. No. 3.2;  
Matches 22; Conservative 11; Mismatches 21; Indels 28; Gaps 3;  
QY 9 RLKGVKVRKCN-----TDDTIGDLKLLIAAQTGRNKKIV----- 43  
DB 20 QLAKPVKINCTRPNNTRKSVRIGPGQTFYATGDIIGIRQACNVSGTWNETLQKVAA 79  
QY 44 -LKKWY--TIFKDHVSLGDYEI 62  
DB 80 QLRKHFNKTIIFASPSGGDVEI 101

RESULT 11

S60524

envelope polyprotein gp41 - human immunodeficiency virus type 1 (isolate CI-45-3) (fragment)

C:Species: human immunodeficiency virus type 1, HIV-1

A:Variety: isolate CI-45-3

C:Date: 20-Jul-1996 #sequence\_revision 13-Mar-1997 #text\_change 26-Aug-1999

C:Accession: S60524

R:Janssens, W.; Heyndrickx, L.; Van de Peer, Y.; Bouckaert, A.; Franssen, K.; Motte, J.;

AIDS 8, 21-26, 1994

A:Title: Molecular phylogeny of part of the env gene of HIV-1 strains isolated in Cote d'Ivoire

A:Reference number: S60521; MUID:94280700

A:Accession: S60524

A:Status: nucleic acid sequence not shown; translation not shown

A:Molecule type: DNA

A:Residues: 1-294 <JAN>

A:Cross-references: EMBL:X72027; NID:g468780; PIDN:CAA50910.1; PID:g468781

A:Experimental source: isolate CI-45-3

A:Note: the nucleotide sequence was submitted to the EMBL Data Library, May 1993

C:Genetics:

A:Gene: env

C:Superfamily: type E retrovirus env polyprotein

C:Keywords: AIDS; glycoprotein; immunodeficiency; polyprotein

Query Match 17.3%; Score 68; DB 2; Length 294;  
Best Local Similarity 26.8%; Pred. No. 3.2;  
Matches 22; Conservative 11; Mismatches 21; Indels 28; Gaps 3;  
QY 9 RLKGVKVRKCN-----TDDTIGDLKLLIAAQTGRNKKIV----- 43  
DB 20 QLAKPVKINCTRPNNTRKSVRIGPGQTFYATGDIIGIRQACNVSGTWNETLQKVAA 79  
QY 44 -LKKWY--TIFKDHVSLGDYEI 62  
DB 80 QLRKHFNKTIIFASPSGGDVEI 101

RESULT 12

S60550

envelope polyprotein gp41 - human immunodeficiency virus type 1 (isolate CI-326-4) (fragment)

C:Species: human immunodeficiency virus type 1, HIV-1

A:Variety: isolate CI-326-4

C:Date: 20-Jul-1996 #sequence\_revision 13-Mar-1997 #text\_change 26-Aug-1999

C:Accession: S60550

R:Janssens, W.; Heyndrickx, L.; Van de Peer, Y.; Bouckaert, A.; Franssen, K.; Motte, J.;

AIDS 8, 21-26, 1994

A:Title: Molecular phylogeny of part of the env gene of HIV-1 strains isolated in Cote d'Ivoire

A:Reference number: S60521; MUID:94280700

A:Accession: S60550

A:Status: nucleic acid sequence not shown; translation not shown

A:Molecule type: DNA

A:Residues: 1-303 <JAN>

A:Cross-references: EMBL:X72054; NID:g468681; PIDN:CAA50935.1; PID:g468682

A:Experimental source: isolate CI-326-4

A:Note: the nucleotide sequence was submitted to the EMBL Data Library, May 1993

C:Genetics:

A:Gene: env

C:Superfamily: type E retrovirus env polyprotein

C:Keywords: AIDS; glycoprotein; immunodeficiency; polyprotein

Query Match 16.8%; Score 66; DB 2; Length 303;  
Best Local Similarity 26.4%; Pred. No. 5.6;  
Matches 23; Conservative 10; Mismatches 20; Indels 34; Gaps 4;  
QY 9 RLKGVKVRKCN-----NTDDTIGDLKLLIAAQTGRNKKIV-- 43  
DB 20 QLSKPVRIINCTRPNNTRKSWRHIGPGRAFYTDDTIGDIRKAYCNVSKTVKMDTLKQ 79  
QY 44 ----LKKWYT----IPKDHVSLGDYEI 62  
DB 80 VARQLRKYNNTIIFANH-SGGDVEI 105

RESULT 13

S60549

envelope polyprotein gp41 - human immunodeficiency virus type 1 (isolate CI-326-3) (fragment)

C:Species: human immunodeficiency virus type 1, HIV-1

A:Variety: isolate CI-326-3

C:Date: 20-Jul-1996 #sequence\_revision 13-Mar-1997 #text\_change 26-Aug-1999

C:Accession: S60549

R:Janssens, W.; Heyndrickx, L.; Van de Peer, Y.; Bouckaert, A.; Franssen, K.; Motte, J.;

AIDS 8, 21-26, 1994

A:Title: Molecular phylogeny of part of the env gene of HIV-1 strains isolated in Cote d'Ivoire

A:Reference number: S60521; MUID:94280700

A:Accession: S60549

A:Status: nucleic acid sequence not shown; translation not shown

A:Molecule type: DNA

A:Residues: 1-303 <JAN>

A:Cross-references: EMBL:X72053; NID:g468679; PIDN:CAA50934.1; PID:g468680

A:Experimental source: isolate CI-326-3

A:Note: the nucleotide sequence was submitted to the EMBL Data Library, May 1993

C:Genetics:

A:Gene: env

C:Superfamily: type E retrovirus env polyprotein

C:Keywords: AIDS; glycoprotein; immunodeficiency; polyprotein

Query Match 16.8%; Score 66; DB 2; Length 303;  
Best Local Similarity 26.4%; Pred. No. 5.6;  
Matches 23; Conservative 10; Mismatches 20; Indels 34; Gaps 4;  
QY 9 RLKGVKVRKCN-----NTDDTIGDLKLLIAAQTGRNKKIV-- 43  
DB 20 QLSKPVRIINCTRPNNTRKSWRHIGPGRAFYTDDTIGDIRKAYCNVSKTVKMDTLKQ 79  
QY 44 ----LKKWYT----IPKDHVSLGDYEI 62  
DB 80 VARQLRKYNNTIIFANH-SGGDVEI 105

RESULT 14

T52254

cytochrome P450 [imported] - Arabidopsis thaliana

C:Species: Arabidopsis thaliana (mouse-ear cress)

C:Date: 20-Oct-2000 #sequence\_revision 20-Oct-2000 #text\_change 01-Dec-2000

C:Accession: T52254

R:Maughan, J.A.; Nugent, J.H.A.; Hallahan, D.L.

submitted to the EMBL Data Library, May 1996

A:Description: Expression of CYP1B7 a cytochrome P450 expressed sequence tag from Arabidopsis thaliana

A:Reference number: Z26004

A:Accession: T52254

A:Status: preliminary; translated from GB/EMBL/DDBT

A:Molecule type: mRNA

A:Residues: 1-504 <MAU>

A:Cross-references: EMBL:X97864; PIDN:CAA66458.1



Query Match	16.6%	Score 65.5;	DB 2;	Length 797;
Best Local Similarity	29.6%;	Pred. No. 17;		
Matches	24;	Conservative 14;	Mismatches 22;	Indels 21; Gaps 5;
QY 4	WCNDRIG--KKYVRKCTDD--TIGDLKKLIAAQGTGR---WNKIVLKKWYTFRDHVS 56			
	: :   : : :           :   :   :   :   :   :   :   :   :   :			
Db 74	IIANSVIGWQETRAKQTLDKLAIQAKPLVRRQSGTRTRSTNEVVL-----DIIIE 126			
QY 57	LG-----DYETHDCMNLEL 70			
	: :   :   :   :   :   :   :   :   :   :   :   :   :   :   :			
Db 127	LGPGDQVWVDGEVVEENLEI 147			

Search completed: June 13, 2001, 20:27:07  
Job time: 344 sec

GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: June 13, 2001, 18:35:07 ; Search time 1697.79 Seconds  
(without alignments)  
1759.785 Million cell updates/sec

Title: US-09-331-930A-1  
Perfect score: 342  
Sequence: 1 gttccaggagattacagctc.....attttttgttaagtgtact 342

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

rchd: 9623517 seqs, 4368049070 residues  
Total number of hits satisfying chosen parameters: 19247034

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database :

EST:

- 1: gb\_estl1:\*
- 2: gb\_estl2:\*
- 3: gb\_estl3:\*
- 4: gb\_estl4:\*
- 5: gb\_estl5:\*
- 6: gb\_estl6:\*
- 7: gb\_estl7:\*
- 8: gb\_estl8:\*
- 9: gb\_estl9:\*
- 10: gb\_estl10:\*
- 11: gb\_estl11:\*
- 12: gb\_estl12:\*
- 13: gb\_estl13:\*
- 14: gb\_estl14:\*
- 15: gb\_estl15:\*
- 16: gb\_estl16:\*
- 17: gb\_estl17:\*
- 18: gb\_estl18:\*
- 19: gb\_estl19:\*
- 20: gb\_estl20:\*
- 21: gb\_estl21:\*
- 22: gb\_estl22:\*
- 23: gb\_estl23:\*
- 24: gb\_estl24:\*
- 25: gb\_estl33:\*
- 26: gb\_estl34:\*
- 27: gb\_estl35:\*
- 28: gb\_estl36:\*
- 29: gb\_estl37:\*
- 30: gb\_estl38:\*
- 31: gb\_estl39:\*
- 32: gb\_estl40:\*
- 33: em\_estba:\*
- 34: em\_estfun:\*
- 35: em\_esthum1:\*
- 36: em\_esthum2:\*
- 37: em\_esthum3:\*
- 38: em\_esthum4:\*
- 39: em\_esthum5:\*
- 40: em\_esthum6:\*
- 41: em\_esthum7:\*
- 42: em\_esthum8:\*
- 43: em\_esthum9:\*

- 44: em\_esthum10:\*
- 45: em\_esthum11:\*
- 46: em\_esthum12:\*
- 47: em\_esthum13:\*
- 48: em\_esthum14:\*
- 49: em\_esthum15:\*
- 50: em\_esthum16:\*
- 51: em\_esthum17:\*
- 52: em\_esthum18:\*
- 53: em\_esthum19:\*
- 54: em\_esthum20:\*
- 55: em\_esthum21:\*
- 56: em\_esthum22:\*
- 57: em\_esthum23:\*
- 58: em\_esthum24:\*
- 59: em\_esthum25:\*
- 60: em\_esthum26:\*
- 61: em\_esthum27:\*
- 62: em\_esthum28:\*
- 63: em\_estin1:\*
- 64: em\_estin2:\*
- 65: em\_estin3:\*
- 66: em\_estin4:\*
- 67: em\_estin5:\*
- 68: em\_estom1:\*
- 69: em\_estom2:\*
- 70: em\_estov1:\*
- 71: em\_estov2:\*
- 72: em\_estpl1:\*
- 73: em\_estpl2:\*
- 74: em\_estpl3:\*
- 75: em\_estpl4:\*
- 76: em\_estpl5:\*
- 77: em\_estpl6:\*
- 78: em\_estpl7:\*
- 79: em\_estpl8:\*
- 80: em\_estpl9:\*
- 81: em\_estpl10:\*
- 82: em\_estro1:\*
- 83: em\_estro2:\*
- 84: em\_estro3:\*
- 85: em\_estro4:\*
- 86: em\_estro5:\*
- 87: em\_estro6:\*
- 88: em\_estro7:\*
- 89: em\_estro8:\*
- 90: em\_estro9:\*
- 91: em\_estro10:\*
- 92: em\_estro11:\*
- 93: em\_estro12:\*
- 94: em\_estro13:\*
- 95: em\_estro14:\*
- 96: em\_estro15:\*
- 97: em\_estro16:\*
- 98: em\_estro17:\*
- 99: em\_estro18:\*
- 100: em\_estro19:\*
- 101: em\_estro20:\*
- 102: gb\_est25:\*
- 103: gb\_est26:\*
- 104: gb\_est27:\*
- 105: gb\_est28:\*
- 106: gb\_est29:\*
- 107: gb\_est30:\*
- 108: gb\_est31:\*
- 109: gb\_est32:\*
- 110: gb\_est41:\*
- 111: gb\_est42:\*
- 112: gb\_est43:\*
- 113: gb\_est44:\*
- 114: gb\_est45:\*
- 115: gb\_est46:\*
- 116: gb\_est47:\*

117: gb\_est48:\*  
118: gb\_est49:\*  
119: gb\_est50:\*  
120: gb\_est51:\*  
121: gb\_est52:\*  
122: gb\_est53:\*  
123: gb\_est54:\*  
124: gb\_est55:\*  
125: gb\_est56:\*  
126: gb\_est57:\*  
127: gb\_est58:\*  
128: gb\_est59:\*  
129: gb\_est60:\*  
130: gb\_est61:\*  
131: gb\_est62:\*  
132: gb\_est63:\*  
133: gb\_est64:\*  
134: gb\_est65:\*  
135: gb\_est66:\*  
136: gb\_est67:\*  
137: gb\_est68:\*  
138: gb\_est69:\*  
139: gb\_est70:\*  
140: gb\_est71:\*  
141: gb\_est72:\*  
142: gb\_est73:\*  
143: gb\_est74:\*  
144: gb\_est75:\*  
145: gb\_est76:\*  
146: gb\_est77:\*  
147: gb\_est78:\*  
148: gb\_est79:\*  
149: gb\_est80:\*  
150: gb\_est81:\*  
151: gb\_est82:\*  
152: gb\_est83:\*  
153: gb\_est84:\*  
154: gb\_est85:\*  
155: gb\_est86:\*  
156: gb\_est87:\*  
157: gb\_est88:\*  
158: gb\_est89:\*  
159: gb\_est90:\*  
160: gb\_est91:\*  
161: gb\_est92:\*  
162: gb\_est93:\*  
163: gb\_est94:\*  
164: gb\_est95:\*  
165: gb\_est96:\*  
166: gb\_est97:\*  
167: gb\_est98:\*  
168: gb\_est99:\*  
169: gb\_est100:\*  
170: gb\_est101:\*  
171: gb\_est102:\*  
172: gb\_est103:\*  
173: gb\_est104:\*  
174: gb\_est105:\*  
175: gb\_est106:\*  
176: gb\_est107:\*  
177: gb\_est108:\*  
178: gb\_est109:\*  
179: gb\_est110:\*  
180: gb\_est111:\*  
181: gb\_est112:\*  
182: gb\_est113:\*  
183: gb\_est114:\*  
184: gb\_est115:\*  
185: gb\_est116:\*  
186: gb\_est117:\*  
187: gb\_est118:\*  
188: gb\_est119:\*  
189: gb\_est120:\*

190: em\_gss\_pln1:\*  
191: em\_gss\_pln2:\*  
192: em\_gss\_pro:\*  
193: em\_gss\_rod1:\*  
194: em\_gss\_rod2:\*  
195: em\_gss\_rod3:\*  
196: em\_gss\_rod4:\*  
197: em\_gss\_rod5:\*  
198: em\_gss\_vrt1:\*  
199: em\_gss\_vrt2:\*  
200: em\_gss\_vrt3:\*  
201: gb\_gss1:\*  
202: gb\_gss2:\*  
203: gb\_gss3:\*  
204: gb\_gss4:\*  
205: gb\_gss5:\*  
206: gb\_gss6:\*  
207: gb\_gss7:\*  
208: gb\_gss8:\*  
209: gb\_gss9:\*  
210: gb\_gss10:\*  
211: gb\_gss11:\*  
212: gb\_gss12:\*  
213: gb\_gss13:\*  
214: gb\_gss14:\*  
215: gb\_gss15:\*  
216: gb\_gss16:\*  
217: gb\_gss17:\*  
218: gb\_gss18:\*  
219: gb\_gss19:\*  
220: gb\_gss20:\*  
221: gb\_gss21:\*  
222: gb\_gss22:\*  
223: gb\_gss23:\*  
224: gb\_gss24:\*  
225: gb\_gss25:\*  
226: gb\_gss26:\*  
227: gb\_gss27:\*  
228: gb\_gss28:\*  
229: gb\_gss29:\*  
230: gb\_gss30:\*  
231: gb\_gss31:\*  
232: gb\_gss32:\*  
233: gb\_gss33:\*  
234: gb\_gss34:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match %	Length	DB	ID	Description
1	254.8	74.5	382	4	AA241830	AA241830 mx12b09.r
2	254.8	74.5	384	137	BE626037	BE626037 ut81c03.x
3	254.8	74.5	392	4	AA265956	AA265956 mz72e08.r
4	254.8	74.5	392	4	AA275337	AA275337 vc07d10.r
5	254.8	74.5	394	20	AI463272	AI463272 uc44f05.x
6	254.8	74.5	395	142	BF017894	BF017894 uw84e06.x
7	254.8	74.5	397	158	W17970	W17970 mb82b05.r1
8	254.8	74.5	398	137	BE628007	BE628007 uu26g05.x
9	254.8	74.5	398	159	W85207	W85207 mf50g10.r1
10	254.8	74.5	398	164	BE197360	BE197360 ug69c04.x
11	254.8	74.5	399	167	BE448243	BE448243 ut81c03.y
12	254.8	74.5	403	1	AA050441	AA050441 mj12d12.r
13	254.8	74.5	406	7	AA462867	AA462867 vf91h07.r
14	254.8	74.5	410	158	W83387	W83387 mf28c03.r1
15	254.8	74.5	411	158	W11593	W11593 ma90a06.r1
16	254.8	74.5	421	158	W11035	W11035 ma76d01.r1
17	254.8	74.5	421	158	W57187	W57187 md67g03.r1
18	254.8	74.5	422	7	AA435358	AA435358 vei5e05.r





JOURNAL  
 COMMENT  
 Unpublished (1996)  
 Contact: Maria M/Mouse EST Project  
 WashU-HHMI Mouse EST Project  
 Washington University School of MedicineP  
 4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108  
 Tel: 314 286 1800  
 Fax: 314 286 1810  
 Email: mouseest@watson.wustl.edu  
 This clone is available royalty-free through LLNL ; contact the  
 IMAGE Consortium (info@image.llnl.gov) for further information.  
 MGI:466731  
 Seq primer: -28ml3 rev2 ET from Amersham.

FEATURES  
 source  
 1..392  
 /organism="Mus musculus"  
 /strain="C57BL/6J"  
 /db\_xref="taxon:10090"  
 /clone="IMAGE:765811"  
 /clone\_lib="Soares mouse lymph node NBLN"  
 /sex="male"  
 /tissue\_type="lymph node"  
 /dev\_stage="4 weeks"  
 /lab\_host="DH10B"  
 /note="Organ: lymph node; Vector: pT7T3D-Pac (Pharmacia)  
 with a modified polylinker; Site\_1: Not I; Site\_2: Eco RI;  
 1st strand cDNA was primed with a Not I - oligo(dT) primer  
 15'  
 TGTACCAATGTAAGTGGGAGCGCGCGTACTTTTTTTTTTTTTTTTTTTTTT  
 3'; double-stranded cDNA was ligated to Eco RI adaptors  
 (Pharmacia), digested with Not I and cloned into the Not I  
 and Eco RI sites of the modified pT7T3 vector. RNA  
 provided by Dr. Bertrand Jordan. Library constructed and  
 normalized by Bento Soares and M. Fatima Bonaldo."  
 BASE COUNT 92 a 102 c 98 g 100 t  
 ORIGIN

Query Match 74.5%; Score 254.8; DB 4; Length 392;  
 Best Local Similarity 89.3%; Pred. No. 4.2e-68;  
 Matches 300; Conservative 0; Mismatches 27; Indels 9; Gaps 2;  
 QY 1 gttccaggagattacagctccagccacaaatgattgagtggtttgcaacgacgcgtctagg 60  
 Db 54 GTTCGAGGATATTGGAGTCCAGCCACCAATGATTGAGGTGGTTTGCAAGACCGCTCTCGG 113  
 QY 61 aaagaaagtcgcggttaagtgaacacccatgacacccatcgaggactgaagaaactgat 120  
 Db 114 AAAGAAAGTCCGCGTTAAGTGAACACCCGATGACACCATCGCGGACTTGAAGAAACTGAT 173  
 QY 121 agcgcccaaaactggcactcgttgaataagatcgttcttaaaaagtggtacacgatttt 180  
 Db 174 ACCTGCTCAAACTGGCACCCCGCTGGAACAAGATCGTCTCTAAAGAGTGTACACGATTTT 233  
 QY 181 taagaccatgtatctctgggagattatgaataccacagatgggatgaacctggagcttta 240  
 Db 234 TTAAGACCAACGCTGCTCTGGGAGATTATGAATACACGATGGGATGAACCTGGAGCTTTA 293  
 QY 241 ttaccagtagagggaattctccacct-----tgcccaacctgttcttccctccc 292  
 Db 294 TTACCAGTAGAGGGGATTCCTCTCCCTCGCCCTGCTCTGCCCTGCCCTCTCTCTCC 353  
 QY 293 atggctcatttaacactgtttagatgctcattttt 328  
 Db 354 AT-CTCATCTGACACTGGTGTAGATGGTCATTTTT 388

RESULT 5  
 AI463272/c  
 LOCUS  
 DEFINITION uc44f05.x1 Soares\_mammary\_gland\_NMLMG Mus musculus cDNA clone  
 IMAGE:I400865 3' similar to WP:F46F11.4 CE10602 ;, mRNA sequence.  
 ACCESSION AI463272  
 VERSION AI463272.1 GI:4317302

EST.  
 house mouse.  
 Mus musculus  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 1 (bases 1 to 394)  
 NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.  
 National Cancer Institute, Cancer Genome Anatomy Project (CGAP),  
 Tumor Gene Index  
 Unpublished (1997)  
 Contact: Robert Strausberg, Ph.D.  
 Tel: (301) 496-1550  
 Email: Robert\_Strausberg@nih.gov  
 This clone is available royalty-free through LLNL ; contact the  
 IMAGE Consortium (info@image.llnl.gov) for further information.  
 MGI:912581  
 This clone was previously sequenced on the 5' end only, this new  
 data is from the 3' end  
 High quality sequence stop: 376.  
 Location/Qualifiers  
 1..394  
 /organism="Mus musculus"  
 /db\_xref="taxon:10090"  
 /clone="IMAGE:1400865"  
 /clone\_lib="Soares\_mammary\_gland\_NMLMG"  
 /sex="female (lactating)"  
 /tissue\_type="mammary gland"  
 /lab\_host="DH10B"  
 /note="Vector: pT7T3D-Pac (Pharmacia) with a modified  
 polylinker; 1st strand cDNA was prepared from mammary  
 gland tissue from a lactating female, and was then primed  
 with a Not I - oligo(dT) primer. Double-stranded cDNA was  
 ligated to Eco RI adaptors (Pharmacia), digested with Not  
 I and cloned into the Not I and Eco RI sites of the  
 modified pT7T3 vector. Library is normalized. Library  
 was constructed by Bento Soares and M. Fatima Bonaldo."  
 BASE COUNT 101 a 95 c 96 g 102 t  
 ORIGIN

Query Match 74.5%; Score 254.8; DB 20; Length 394;  
 Best Local Similarity 89.3%; Pred. No. 4.2e-68;  
 Matches 300; Conservative 0; Mismatches 27; Indels 9; Gaps 2;  
 QY 1 gttccaggagattacagctccagccacaaatgattgagtggtttgcaacgacgcgtctagg 60  
 Db 370 GTTCGAGGATATTGGAGTCCAGCCACCAATGATTGAGGTGGTTTGCAACACCGCTCTCGG 311  
 QY 61 aaagaaagtcgcggttaagtgaacacccatgacacccatcgaggactgaagaaactgat 120  
 Db 310 AAAGAAAGTCCGCGTTAAGTGAACAACACCGATGACACCATCGCGGACTTGAAGAACTGAT 251  
 QY 121 agcgcccaaaactggcactcgttgggaataagatcgttcttaaaaagtggtacacgatttt 180  
 Db 250 AGCTGCTCAAACTGGCACCCCGCTGGAACAAGATCGTCTCTTAAAGAGTGGTACACGATTTT 191  
 QY 181 taagaccatgtatctctgggagattatgaataccacagatgggatgaacctggagcttta 240  
 Db 190 TTAAGACCAACGCTGCTCTGGGAGATTATGAATACACGATGGGATGAACCTGGAGCTTTA 131  
 QY 241 ttaccagtagagggaattctccacct-----tgcccaacctgttcttccctccc 292  
 Db 130 TTACCAGTAGAGGGGATTCCTCTCTCTCGCCCTGCTCTGCCCTGCCCTCTCTCTCC 71  
 QY 293 atggctcatttaacactgtttagatgctcattttt 328  
 Db 70 AT-CTCATCTGACACTGGTGTAGATGGTCATTTTT 36

RESULT 6  
 BF017894/c  
 LOCUS  
 DEFINITION uw84e06.x1 Soares\_mammary\_gland\_NMLMG Mus musculus cDNA clone

[illegible]

QY 241 ttaccagtagaggggaattctccacct-----tgcccaaccttgctttctctctccc 292  
 Db 270 TTACCAGTAGAGGGGATTCCTCTCTCCCGCCTCTGCGCCCTGCCCTCTCTCC 329  
 QY 293 atggctcatttaaacactgtgttagatgctcatttt 328  
 Db 330 AT-CCTCATCTGACACTGGTGTAGATGTCATTTT 364

RESULT 8  
 LOCUS BE628007/c 398 bp mRNA EST 24-AUG-2000  
 DEFINITION uu26g05.x1 Soares\_mammary\_gland\_NMLMG Mus musculus cDNA clone  
 IMAGE:3373112 3' similar to WP:F46F11.4 CE10602 ; mRNA sequence.  
 ACCESSION BE628007  
 VERSION BE628007.1 GI:9908452  
 KEYWORDS EST.  
 SOURCE house mouse.  
 ORGANISM Mus musculus  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 REFERENCE 1 (bases 1 to 398)  
 AUTHORS NCI-CCAP <http://www.ncbi.nlm.nih.gov/ncicgap>.  
 TITLE National Cancer Institute, Cancer Genome Anatomy Project (CGAP),  
 Tumor Gene Index  
 JOURNAL Unpublished (1997)  
 COMMENT Contact: Robert Strausberg, Ph.D.  
 Tel: (301) 496-1550  
 Email: Robert.Strausberg@nih.gov  
 This clone is available royalty-free through LLNL ; contact the  
 IMAGE Consortium ([info@image.llnl.gov](mailto:info@image.llnl.gov)) for further information.  
 MGI:1082716.

FEATURES  
 source  
 1. 398  
 /organism="Mus musculus"  
 /db\_xref="taxon:10090"  
 /clone="IMAGE:3373112"  
 /clone\_lib="Soares\_mammary\_gland\_NMLMG"  
 /sex="female (lactating)"  
 /tissue\_type="mammary gland"  
 /lab\_host="DH10B"  
 /note="Vector: p7T3D-Pac (Pharmacia) with a modified  
 polylinker; 1st strand cDNA was prepared from mammary  
 gland tissue from a lactating female, and was then primed  
 with a Not I - oligo(dT) primer. Double-stranded cDNA was  
 ligated to Eco RI adaptors (Pharmacia), digested with Not  
 I and cloned into the Not I and Eco RI sites of the  
 modified p7T3 vector. Library is normalized. Library  
 was constructed by Bento Soares and M. Fatima Bonaldo."  
 102 a 96 c 99 g 101 t

Query Match 74.5%; Score 254.8; DB 137; Length 398;  
 Best Local Similarity 89.3%; Pred. No. 4.3e-68;  
 Matches 300; Conservative 0; Mismatches 27; Indels 9; Gaps 2;

QY 1 gtccagagattacagctccagccacacatgattgagttggtttcaacagcagctctagg 60  
 Db 370 GTTCGAGGATATTGAGCTCCAGCCACCAATGATTGAGTGGTTTCAACAGCCGCTCGG 311  
 QY 61 aaagaaagtcgcggttaagtcaacaccgatacacaccatcggggacttgaagaaactgat 120  
 Db 310 AAAGAAGATCGCGTTAGTGAACACCCGATGACACCATCGCGACTTGAAGAAACTGAT 251  
 QY 121 agcgccccaaactggcactctgtggaataaagatcttcttaaaaagtgtgacacgatttt 180  
 Db 250 AGCTGCTCAAACTGCGCCCGTGAACAAGATCCTTCTTAAAGAGTGTACACGATTTT 191  
 QY 181 taaggaccatgtatctctgggagattatgaatcacacatgggagctggaagcttta 240  
 Db 190 TAAGGACCACGTGCTCTGGGAGATTATGAATCCACGATGGGATGAACCTGGAGCTTTA 131

QY 241 ttaccagtagaggggaattctccacct-----tgcccaaccttgctttctctctccc 292  
 Db 130 TTACCAGTAGAGGGGATTCCTCTCTCCCGCCTCTGCGCCCTGCCCTCTCTCC 71  
 QY 293 atggctcatttaaacactgtgttagatgctcatttt 328  
 Db 70 AT-CCTCATCTGACACTGGTGTAGATGTCATTTT 36

RESULT 9  
 LOCUS W85207 398 bp mRNA EST 12-SEP-1996  
 DEFINITION m50g10.r1 Soares mouse embryo NbME13.5 14.5 Mus musculus cDNA  
 clone IMAGE:408546 5', mRNA sequence.  
 ACCESSION W85207  
 VERSION W85207.1 GI:1541533  
 KEYWORDS EST.  
 SOURCE house mouse.  
 ORGANISM Mus musculus  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 REFERENCE 1 (bases 1 to 398)  
 AUTHORS Marra, M., Hillier, L., Allen, M., Bowles, M., Dietrich, N., Dubuque, T.,  
 Geiseli, S., Kucaba, T., Lacy, M., Le, M., Martin, J., Morris, M.,  
 Scheilenberg, K., Steptoe, M., Tan, F., Underwood, K., Moore, B.,  
 Theising, B., Wylie, T., Lennon, G., Soares, B., Wilson, R. and  
 Waterston, R.  
 TITLE The WashU-HMI Mouse EST Project  
 JOURNAL Unpublished (1996)  
 COMMENT On Sep 12, 1996 this sequence version replaced gi:1397677.  
 Contact: Marra M/Mouse EST Project  
 WashU-HMI Mouse EST Project  
 Washington University School of Medicine  
 4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108  
 Tel: 314 286 1800  
 Fax: 314 286 1810  
 Email: mouseest@wustl.edu  
 This clone is available royalty-free through LLNL ; contact the  
 IMAGE Consortium ([info@image.llnl.gov](mailto:info@image.llnl.gov)) for further information.  
 MGI:252314

Seq primer: -28M13 rev2 from Amersham  
 High quality sequence stop: 392.  
 Location/Qualifiers  
 1. 398  
 /organism="Mus musculus"  
 /strain="C57BL/6J"  
 /db\_xref="taxon:10090"  
 /clone="IMAGE:408546"  
 /clone\_lib="Soares mouse embryo NbME13.5 14.5"  
 /sex="unknown"  
 /tissue\_type="embryo"  
 /dev\_stage="13.5-14.5dpc total fetus"  
 /lab\_host="DH10B"  
 /note="Vector: p7T3D-Pac (Pharmacia) with a modified  
 polylinker; Site 1: Not I; Site 2: Eco RI; 1st strand cDNA  
 was primed with a Not I - oligo(dT) primer [5',  
 TGTTCACCAATCTGAAGTGGAGCGCGGAAATTTTGTGTGTGTGTGTGTGTGTGT  
 T 3'], on equal amounts of mRNA from 2 13.5dpc and 2  
 14.5dpc embryos [total RNA provided by Minoru Ko, Wayne  
 State Univ., from 2 ]; double-stranded cDNA was ligated to  
 Eco RI adaptors (Pharmacia), digested with Not I and  
 cloned into the Not I and Eco RI sites of the modified  
 p7T3 vector. Library went through one round of  
 normalization, and was constructed by Bento Soares and  
 M. Fatima Bonaldo."  
 100 a 99 c 97 g 102 t

FEATURES  
 source

Query Match 74.5%; Score 254.8; DB 159; Length 398;  
 Best Local Similarity 89.3%; Pred. No. 4.3e-68;  
 Matches 300; Conservative 0; Mismatches 27; Indels 9; Gaps 2;











GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: June 13, 2001, 18:39:17 ; Search time 86.89 Seconds  
(without alignments)  
687.238 Million cell updates/sec

Title: US-09-331-930A-1

Perfect score: 342

Sequence: 1 gttccaggagattacagctc.....attttttgtaagtgtact 342

Scoring table: IDENTITY\_NUC

Gapop 10.0 , Gapext 1.0

Archived: 302621 seqs, 87301344 residues

Total number of hits satisfying chosen parameters: 605242

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Issued\_Patents\_NA.\*

1: /cgn2.6/ptodata/1/ina/5A\_COMB.seq.\*  
2: /cgn2.6/ptodata/1/ina/5B\_COMB.seq.\*  
3: /cgn2.6/ptodata/1/ina/6A\_COMB.seq.\*  
4: /cgn2.6/ptodata/1/ina/6B\_COMB.seq.\*  
5: /cgn2.6/ptodata/1/ina/PCTUS\_COMB.seq.\*  
6: /cgn2.6/ptodata/1/ina/backfiles1.seq.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	35.4	10.4	12492	6	Patent No. 5206163
2	32.2	9.4	3840	1	Sequence 1, Appli
3	32.2	9.4	3840	3	Sequence 1, Appli
4	32.2	9.4	3840	3	Sequence 1, Appli
5	32.2	9.4	3840	6	Patent No. 5510474
6	31.8	9.3	1757	1	Sequence 49, Appli
7	30.6	8.9	4245	5	Sequence 516, App
8	30	8.8	3571	4	Sequence 132, App
9	30	8.8	3571	4	Sequence 9, Appli
10	29.6	8.7	581	2	Sequence 22, Appli
11	29.6	8.7	581	3	Sequence 22, Appli
12	29	8.5	7218	1	Sequence 14, Appli
13	28.4	8.3	707	2	Sequence 22, Appli
14	28.4	8.3	7218	1	Sequence 14, Appli
15	28.2	8.2	1800	1	Sequence 3, Appli
16	28.2	8.2	2022	2	Sequence 96, Appli
17	28.2	8.2	2022	3	Sequence 96, Appli
18	28.2	8.2	2022	3	Sequence 96, Appli
19	28.2	8.2	2022	5	Sequence 96, Appli
20	28.2	8.2	2022	5	Sequence 96, Appli
21	28.2	8.2	2127	2	Sequence 96, Appli
22	28.2	8.2	2127	3	Sequence 95, Appli
23	28.2	8.2	2127	3	Sequence 95, Appli
24	28.2	8.2	2127	5	Sequence 95, Appli
25	28.2	8.2	2127	5	Sequence 95, Appli
26	28.2	8.2	2252	4	Sequence 13, Appli
27	28	8.2	554	3	Sequence 16, Appli

28 28 8.2 554 3 US-09-360-220-16 Sequence 16, Appli  
29 28 8.2 2241 3 US-08-840-146-1 Sequence 1, Appli  
30 28 8.2 2241 3 US-09-360-220-1 Sequence 1, Appli  
31 27.8 8.1 3621 2 US-09-019-201A-1 Sequence 1, Appli  
32 27.6 8.1 366 4 US-08-881-771A-7 Sequence 7, Appli  
33 27.6 8.1 1096 4 US-08-881-771A-5 Sequence 5, Appli  
34 27.6 8.1 2105 4 US-09-088-425-3 Sequence 3, Appli  
35 27.4 8.0 231 1 US-08-450-834-1 Sequence 1, Appli  
c 37 27.2 8.0 831 1 US-08-770-379-17 Sequence 5, Appli  
c 38 27.2 8.0 35100 2 US-08-757-669A-17 Sequence 17, Appli  
c 39 27.2 7.9 1817 3 US-08-564-496C-14 Sequence 14, Appli  
c 40 27 7.9 3349 2 US-08-920-234-1 Sequence 1, Appli  
c 41 26.6 7.8 2664 2 US-08-942-819-1 Sequence 1, Appli  
c 42 26.6 7.8 285 6 5487983-3 Patent No. 5487983  
c 43 26.6 7.8 4481 4 US-08-867-611-1 Sequence 1, Appli  
c 44 26.6 7.8 4481 5 PCT-US92-06965A-6 Sequence 6, Appli  
c 45 26.6 7.8 5600 4 US-08-867-611-3 Sequence 3, Appli

#### ALIGNMENTS

RESULT 1

5206163-2

; Patent No. 5206163

; APPLICANT: RENARD, ANDRE; DINA, DINO; MARTIAL, JOSEPH

; TITLE OF INVENTION: DNA ENCODING BOVINE DIARRHEA

; VIRUS PROTEIN

; NUMBER OF SEQUENCES: 3

; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US/07/550,816

; FILING DATE: 06-JUL-1990

; PRIOR APPLICATION DATA:

; APPLICATION NUMBER: 331,037

; FILING DATE: 29-MAR-1989

; APPLICATION NUMBER: 752,981

; FILING DATE: 08-JUL-1985

; SEQ ID NO:2

; LENGTH:12492

5206163-2

Query Match 10.4%; Score 35.4; DB 6; Length 12492;  
Best Local Similarity 45.8%; Pred. No. 0.017; Indels 0; Gaps 0;  
Matches 111; Conservative 0; Mismatches 126;

Qy 28 aatgattgaggtgtttgcaacgacctgtctaggaagaaagtcgcttaagtgaacac 87  
Db 5148 aaaaatgcagatcttcgtgaaacccctgacccgcaagaccatcacccctggaggtgagcc 5207  
Qy 88 cgatgacaccatcggtggtgaaagaaactgatacgcccaaaactggcactcgttgaa 147  
Db 5208 cagtgcaccatcagaaagctgaaagccagatcaggaaggaagcctccctcctga 5267  
Qy 148 taagatcgtttctaaagagtggtacacatttttaagaccatgtatctctggagatt 207  
Db 5268 ccagcagaggtctcttttgcggcagcagctggaagagtcgctcctctcttctgatt 5327  
Qy 208 tgaatccacatgggtgagccttgagctttattaccagtagaggggaattcctcc 264  
Db 5328 caacatccagaagagtcgacccctgcacctcgtctcgtctgaggggtagtgagcc 5384

RESULT 2

US-08-462-092-1

; Sequence 1, Application US/08462092

; Patent No. 5614399

; GENERAL INFORMATION:

; APPLICANT: Peter H. Quail

; APPLICANT: Alan H. Christensen

; APPLICANT: Howard P. Hershey

; APPLICANT: Robert A. Sharrock

CORRESPONDENCE ADDRESS:  
ADDRESSEE: Jeff Lloyd  
STREET: 2421 N.W. 41st Street, Suite A-1  
CITY: Gainesville  
STATE: FL  
COUNTRY: USA  
ZIP: 32606  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC Compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/746,822  
FILING DATE:  
CLASSIFICATION:  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US/08/462,092  
FILING DATE: 5-JUNE-1995  
ATTORNEY/AGENT INFORMATION:  
NAME: Lloyd, Jeff  
REGISTRATION NUMBER: 35,589  
REFERENCE/DOCKET NUMBER: 08/462,092  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 904-375-8100  
TELEFAX: 904-372-5800  
INFORMATION FOR SEQ ID NO: 1:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 3840 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: DNA (genomic)  
MOLECULE TYPE: DNA (genomic)  
FEATURE:  
NAME/KEY: CDS  
LOCATION: 1993..3591  
US-08-746-822-1

```

Query Match          9.4%; Score 32.2; DB 3; Length 3840;
Best Local Similarity 47.7%; Pred. No. 0.13;
Matches 94; Conservative 0; Mismatches 103; Indels 0; Gaps 0;

.
QY 59 ggaagaagtcgcgtaagtgcacacaccgatgcacaccatcggggacttgaagaactg 118
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 3160 GGCAAGACCAATCACCCTGGAGGTGGAGAGCTCTGACACCATGTGCAAGGCCAAG 3219
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 119 atagcggcccaaaactggcactcgttggataagatcgtctctaaaaaagtgtcacacgatt 178
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 3220 ATCCAGGACAAAGGAGGCGATTCCTCCCGACGACGAGCGTCTGATCTTTGCGGGCAAGCAG 3279
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 179 tttaaggacacatgtatctctcggggattatgaataccacgatgggatgaacctgagctt 238
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 3280 CTGGAGGATGCCCGCACTCTCCGGACTTACAACATCCAGAGGAGGAGCACCCTTCACCTT 3339
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 239 tattaccagttagaggg 255
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 3340 GTTCTCGGCTCAGAGG 3356
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

RESULT 4
US-09-094-350-1
; Sequence 1, Application US/09094350
; Patent No. 6054574
; GENERAL INFORMATION:
; APPLICANT: Peter H. Quail
; APPLICANT: Alan H. Christensen
; APPLICANT: Howard P. Hershey
; APPLICANT: Robert A. Sharrock
; APPLICANT: Thomas D. Sullivan
; TITLE OF INVENTION: PLANT UBQUITIN PROMOTER SYSTEM
; NUMBER OF SEQUENCES: 2

```

DECEMBER 1987



[illegible]

```

; Patent No. 6136603
; GENERAL INFORMATION:
; APPLICANT: Dean, Nicholas M.
; APPLICANT: Karras, James G
; APPLICANT: McKay, Robert
; TITLE OF INVENTION: ANTISENSE MODULATION OF INTERLEUKIN-5 SIGNAL
; TITLE OF INVENTION: TRANSDUCTION
; FILE REFERENCE: ISPH-0340
; CURRENT APPLICATION NUMBER: US/09/280,799
; CURRENT FILING DATE: 1999-03-26
; NUMBER OF SEQ ID NOS: 208
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 132
; LENGTH: 3571
; TYPE: DNA
; ORGANISM: Mus musculus
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence:Synthetic
US-09-280-799-132

Query Match      8.8%; Score 30; DB 4; Length 3571;
Best Local Similarity 49.4%; Pred. No. 0.78;
Matches 78; Conservative 0; Mismatches 80; Indels 0; Gaps 0

Qy 176 atttttaaggaccatgtatctctctggagattatgaatccacagatgggagtaacctggag 235
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 2600 AATTTTTTAAGACCAATTTCTTCAGTGAACCTAGAGATAATTAATTTGGCTAAACTGGTTG 2541
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Ov 236 ctttattaccagtagagggaaatttctccacgtttaccacaacctttactttctctccacatg 295

```



```

; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Maki, David J.
; REGISTRATION NUMBER: 31,392
; REFERENCE/DOCKET NUMBER: 210121.422
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (206) 622-4900
; TELEFAX: (206) 682-6031
; INFORMATION FOR SEQ ID NO: 22:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 581 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; US-08-557-309B-22

Query Match      8.7%; Score 29.6; DB 2; Length 581;
Best Local Similarity 47.8%; Pred. No. 0.37; Indels 0; Gaps 0;
Matches 86; Conservative 0; Mismatches 94; Indels 0; Gaps 0;

QY 59 ggaagaagtcgcgcttaagtgcacacccgacacccatcgggacgttgaaagaaactg 118
    ||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 72 GGCAAGACGATCGCGCTCGAGGTGGATCCAGCGACACCATTTGAGACCTGAAGCGGAAG 131
    ||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 119 atagcgcccaaaactggcactcgttggaataagatcgttcttaaaaagtggtacacgatt 178
    ||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 132 ATCCAGGACAAAGGAGGCGCATTCGCGCGACGACGCGCTGATCTTCGCTGGCAAGCAG 191
    ||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 179 ttaaggaccatgtatctctggagattatgaatccacgattgggatgaacctggagctt 238
    ||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 192 CTGGAGGACGCGCGACGCTCGCAGACTACAAACATCCAGAGGAGTCCACGCTGCACCTT 251

RESULT 11
US-08-834-306-22
; Sequence 22, Application US/08834306
; Patent No. 6054135
; GENERAL INFORMATION:
; APPLICANT: Reed, Steven G.
; APPLICANT: Skeiky, Yasir A.W.
; APPLICANT: Lodes, Michael J.
; APPLICANT: Houghton, Raymond L.
; TITLE OF INVENTION: COMPOUNDS AND METHODS FOR THE DETECTION AND PREVENTION OF
; NUMBER OF SEQUENCES: 65
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: SEED and BERRY LLP
; STREET: 6300 Columbia Center, 701 Fifth Avenue
; CITY: Seattle
; STATE: Washington
; COUNTRY: USA
; ZIP: 98104-7092
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/834,306
; FILING DATE: 15-APR-1997
; CLASSIFICATION: 424
; ATTORNEY/AGENT INFORMATION:
; NAME: Maki, David J.
; REGISTRATION NUMBER: 31,392
; REFERENCE/DOCKET NUMBER: 210121.422C1
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (206) 622-4900
; TELEFAX: (206) 682-6031
; INFORMATION FOR SEQ ID NO: 22:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 581 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single

```

```

; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Maki, David J.
; REGISTRATION NUMBER: 31,392
; REFERENCE/DOCKET NUMBER: 210121.422
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (206) 622-4900
; TELEFAX: (206) 682-6031
; INFORMATION FOR SEQ ID NO: 22:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 581 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; US-08-557-309B-22

Query Match      8.7%; Score 29.6; DB 2; Length 581;
Best Local Similarity 47.8%; Pred. No. 0.37; Indels 0; Gaps 0;
Matches 86; Conservative 0; Mismatches 94; Indels 0; Gaps 0;

QY 59 ggaagaagtcgcgcttaagtgcacacccgacacccatcgggacgttgaaagaaactg 118
    ||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 72 GGCAAGACGATCGCGCTCGAGGTGGATCCAGCGACACCATTTGAGACCTGAAGCGGAAG 131
    ||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 119 atagcgcccaaaactggcactcgttggaataagatcgttcttaaaaagtggtacacgatt 178
    ||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 132 ATCCAGGACAAAGGAGGCGCATTCGCGCGACGACGCGCTGATCTTCGCTGGCAAGCAG 191
    ||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 179 ttaaggaccatgtatctctggagattatgaatccacgattgggatgaacctggagctt 238
    ||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 192 CTGGAGGACGCGCGACGCTCGCAGACTACAAACATCCAGAGGAGTCCACGCTGCACCTT 251

RESULT 11
US-08-834-306-22
; Sequence 22, Application US/08834306
; Patent No. 6054135
; GENERAL INFORMATION:
; APPLICANT: Reed, Steven G.
; APPLICANT: Skeiky, Yasir A.W.
; APPLICANT: Lodes, Michael J.
; APPLICANT: Houghton, Raymond L.
; TITLE OF INVENTION: COMPOUNDS AND METHODS FOR THE DETECTION AND PREVENTION OF
; NUMBER OF SEQUENCES: 65
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: SEED and BERRY LLP
; STREET: 6300 Columbia Center, 701 Fifth Avenue
; CITY: Seattle
; STATE: Washington
; COUNTRY: USA
; ZIP: 98104-7092
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/834,306
; FILING DATE: 15-APR-1997
; CLASSIFICATION: 424
; ATTORNEY/AGENT INFORMATION:
; NAME: Maki, David J.
; REGISTRATION NUMBER: 31,392
; REFERENCE/DOCKET NUMBER: 210121.422C1
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (206) 622-4900
; TELEFAX: (206) 682-6031
; INFORMATION FOR SEQ ID NO: 22:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 581 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single

```

TOPOLOGY: linear  
US-08-834-306-22

Query Match 8.7%; Score 29.6; DB 3; Length 581;  
Best Local Similarity 47.8%; Pred. No. 0.37;

Matches 86; Conservative 0; Mismatches 94; Indels 0; Gaps 0;

```
Qy 59 ggaagaagtcgcgttaagtcaacacogatgacaccatcggggacttgaagaactg 118
Dy 72 GCAAGACGATCGCTCGAGTGAATCCAGCGCACCATGAGAACGTGAAGCGGAAG 131
Qy 119 atagcgcccaactgcactcgttgaataagatcgttcttaaaagtgtacacgatt 178
Dy 132 ATCCAGGACAGGAGGCGATTCGCCCGACACGAGCGCTGATCTTCGCTGGGACGAG 191
Qy 179 tttaagaccatgtatctcgtggagattatgaatccacatcgatggatgaactgagatt 238
Dy 192 CTGAGGACGCGCCGCTCGCAGACTACAACTCCAGAAGAGTCCACGCTGCACCTT 251
```

## RESULT 12

US-232-463-14/c

Sequence 14, Application US/08232463

Patent No. 5670367

GENERAL INFORMATION:

APPLICANT: DORNER, F.

APPLICANT: SCHEIFLINGER, F.

APPLICANT: FALKNER, F. G.

TITLE OF INVENTION: RECOMBINANT FOWLPOX VIRUS

NUMBER OF SEQUENCES: 52

CORRESPONDENCE ADDRESS:

ADDRESSEE: Foley & Lardner

STREET: 1800 Diagonal Road, Suite 500

CITY: Alexandria

STATE: VA

COUNTRY: USA

ZIP: 22313-0299

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: PatentIn Release #1.0, Version #1.25

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/232.463

FILING DATE:

CLASSIFICATION: 435

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US/07/935,313

FILING DATE:

APPLICATION NUMBER: EP 91 114 300.6

FILING DATE: 26-AUG-1991

ATTORNEY/AGENT INFORMATION:

NAME: BENT, Stephen A.

REGISTRATION NUMBER: 29,768

REFERENCE/DOCKET NUMBER: 30472/114 IMMU

TELECOMMUNICATION INFORMATION:

TELEPHONE: (703)836-9300

TELEFAX: (703)683-4109

TELEX: 899149

INFORMATION FOR SEQ ID NO: 14:

SEQUENCE CHARACTERISTICS:

LENGTH: 7218 base pairs

TYPE: nucleic acid

STRANDEDNESS: single

TOPOLOGY: linear

IMMEDIATE SOURCE:

CLONE: pTZgpt-Fls

US-08-232-463-14

Query Match 8.5%; Score 29; DB 1; Length 7218;  
Best Local Similarity 4.0%; Pred. No. 2.8;

Matches 11; Conservative 146; Mismatches 116; Indels 0; Gaps 0;

```
Qy 6 aggagattacagctccagccacaatgattgagtggtttgcaacgacctcttaggaaaga 65
Dy 1319 RRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1260
Qy 66 aagtcgcgttaagtgcacacccgatgacaccatcggggacttgaagaactgtagcg 125
Dy 1259 RRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1200
Qy 126 cccaaactggcactcgttgggaataagatcgttcttlaaaagtggtacacgatttt 185
Dy 1199 RRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1140
Qy 186 accatgtatctctggagattatgaatccacgatgggatgaacctggagctttattac 245
Dy 1139 RRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1080
Qy 246 agtagaggggaattctccaccttgcacacct 278
Dy 1079 RRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1047
```

## RESULT 13

US-08-726-306A-22

Sequence 22, Application US/08726306A

Patent No. 5958684

GENERAL INFORMATION:

APPLICANT: van Leeuwen, Frederik Willem

APPLICANT: Burbach, Johannes Peter Henri

APPLICANT: Grosveld, Franklin G.

TITLE OF INVENTION: DIAGNOSIS METHOD AND REAGENTS

NUMBER OF SEQUENCES: 189

CORRESPONDENCE ADDRESS:

ADDRESSEE: Banner & Witcoff, Ltd.

STREET: 1 Financial Center

CITY: Boston

STATE: MA

COUNTRY: US

ZIP: 02111

COMPUTER READABLE FORM:

MEDIUM TYPE: Diskette, 3.50 inch, 1.44 Mb storage

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Wordperfect 6.1

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/726,306A

FILING DATE: 02-Oct-1996

PRIOR APPLICATION DATA:

APPLICATION NUMBER: GB 95/20080.4

FILING DATE: 02-Oct-1995

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US 60/009,832

FILING DATE: 01-Jan-1996

ATTORNEY/AGENT INFORMATION:

NAME: Williams, Ph.D., Kathleen M.

REGISTRATION NUMBER: 34,380

REFERENCE/DOCKET NUMBER: 96,048-A (3255/00784)

TELECOMMUNICATION INFORMATION:

TELEPHONE: (617) 345-9100

TELEFAX: (617) 345-9111

INFORMATION FOR SEQ ID NO: 22:

SEQUENCE CHARACTERISTICS:

LENGTH: 707 base pairs

TYPE: nucleic acid

STRANDEDNESS: double

TOPOLOGY: linear

MOLECULE TYPE: cDNA

US-08-726-306A-22

Query Match 8.3%; Score 28.4; DB 2; Length 707;  
Best Local Similarity 46.1%; Pred. No. 1.1;



```

Qy 204 attatgaatccacgatgggataaaccctggagctttattaccagtagaggggaattcctc 263
Db 1209 ATTCTACAGGCAAAATATGCAAAAGGAGCCAAACCTGTAAACCCAGCATTTGGGATGTCAAG 1268
Qy 264 caccttgcccaacccttgcttccctcctcccatggctcatttaaacactgttgtagatgetca 323
Db 1269 ACTGGAGCTAACTCTCATTCAATTCAAAAGTCTTTTATACAATTTCTGTACATACCTT 1328
Qy 324 ttttttggttaagtgtta 340
Db 1329 TTTTTTTTTTAAGAGAA 1345

```

Search completed: June 13, 2001, 20:21:14  
Job time: 6117 sec

GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: June 13, 2001, 18:35:57 ; Search time 1224.15 Seconds  
(without alignments)  
4120.220 Million cell updates/sec

Title: US-09-331-930A-1  
Perfect score: 342  
Sequence: 1 gttccaggagattacagctc.....attttttgtaagtgtact 342

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 1283235 seqs, 7373929652 residues 2566470  
Total number of hits satisfying chosen parameters:

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000  
Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : GenEmbl.\*  
1: gb\_ba1.\*  
2: gb\_ba2.\*  
3: gb\_ba3.\*  
4: gb\_in1.\*  
5: gb\_in2.\*  
6: gb\_in3.\*  
7: gb\_om.\*  
8: gb\_ov.\*  
9: gb\_pat1.\*  
10: gb\_pat2.\*  
11: gb\_ph.\*  
12: gb\_pl1.\*  
13: gb\_pl2.\*  
14: gb\_pl3.\*  
15: gb\_pl4.\*  
16: em\_ba1.\*  
17: em\_ba2.\*  
18: em\_fun.\*  
19: em\_htgo\_hum.\*  
20: em\_htgo\_inv.\*  
21: em\_htgo\_rod.\*  
22: em\_htg\_hum1.\*  
23: em\_htg\_hum2.\*  
24: em\_htg\_hum3.\*  
25: em\_htg\_hum4.\*  
26: em\_htg\_hum5.\*  
27: em\_htg\_hum6.\*  
28: em\_htg\_hum7.\*  
29: em\_htg\_hum8.\*  
30: em\_htg\_inv1.\*  
31: em\_htg\_inv2.\*  
32: em\_htg\_other.\*  
33: em\_htg\_rod.\*  
34: em\_hum1.\*  
35: em\_hum2.\*  
36: em\_hum3.\*  
37: em\_hum4.\*  
38: em\_hum5.\*  
39: em\_hum6.\*  
40: em\_hum7.\*  
41: em\_in.\*  
42: em\_om.\*  
43: em\_or.\*

44: em\_ov.\*  
45: em\_pat.\*  
46: em\_ph.\*  
47: em\_pl.\*  
48: em\_ro.\*  
49: em\_sts.\*  
50: em\_sy.\*  
51: em\_un.\*  
52: em\_vi.\*  
53: gb\_sts1.\*  
54: gb\_sts2.\*  
55: gb\_sts3.\*  
56: gb\_sy.\*  
57: gb\_un.\*  
58: gb\_vil.\*  
59: gb\_vil2.\*  
60: gb\_vil3.\*  
61: gb\_vil4.\*  
62: gb\_vil5.\*  
63: gb\_vil6.\*  
64: gb\_vil7.\*  
65: gb\_vil8.\*  
66: gb\_vil9.\*  
67: gb\_vil10.\*  
68: gb\_vil11.\*  
69: gb\_vil12.\*  
70: gb\_vil13.\*  
71: gb\_vil14.\*  
72: gb\_vil15.\*  
73: gb\_vil16.\*  
74: gb\_vil17.\*  
75: gb\_vil18.\*  
76: gb\_vil19.\*  
77: gb\_vil20.\*  
78: gb\_vil21.\*  
79: gb\_vil22.\*  
80: gb\_vil23.\*  
81: gb\_vil24.\*  
82: gb\_vil25.\*  
83: gb\_vil26.\*  
84: gb\_vil27.\*  
85: gb\_vil28.\*  
86: gb\_vil29.\*  
87: gb\_vil30.\*  
88: gb\_vil31.\*  
89: gb\_vil32.\*  
90: gb\_vil33.\*  
91: gb\_vil34.\*  
92: gb\_vil35.\*  
93: gb\_vil36.\*  
94: gb\_vil37.\*  
95: gb\_vil38.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	DB ID	Description
1	324	94.7	368	94	AF318186 Psammomys
2	234.6	68.6	204394	94	AC079644 Mus muscu
3	232.4	68.0	204394	94	AC079644 Mus muscu
4	195.2	57.1	176560	67	AC021852 Homo sapi
5	195.2	57.1	185380	64	AC015941 Homo sapi
6	195.2	57.1	186809	70	AC026620 Homo sapi
7	125.8	36.8	201602	75	AC074146 Mus muscu
8	125.8	36.8	219217	74	AC069459 Mus muscu
9	121.4	35.5	156342	70	AC026142 Homo sapi
10	121.4	35.5	160426	74	AC068372 Homo sapi
11	118.8	34.7	59762	12	AB023032 Arabidops

```

c 12 111 32.5 88010 13 ATT14D3
c 13 111 32.5 182320 65 AC016808 Homo sapi
c 14 109 31.9 94838 74 AC068782_4
      15 FHU10292
c 16 102.8 30.1 1805 15 YSCPFGS
c 17 102.8 30.1 2201 14 SCYNR032W
c 18 102.8 30.1 3204 14 SCYNR033W
c 19 102.8 30.1 3840 15 YSCPAMIBEN
c 20 97.8 28.6 1065 53 CNS06PDK
c 21 88.4 25.8 2906 6 DVAJ5174
c 22 78.8 23.0 236692 75 AC073773
c 23 78.8 23.0 266229 75 AC073773
c 24 75.6 22.1 139376 93 HS095742
c 25 75.6 22.1 172827 86 AC007216
c 26 75 21.9 26412 66 AC020443
c 27 75 21.9 111455 60 AC007549
c 28 75 21.9 126593 60 AC008340
c 29 70 20.5 871 54 CNS07D9X
c 30 65.8 19.2 64036 72 AC040921
c 31 65.8 19.2 138036 87 AC008752
c 32 63.8 18.7 36532 5 CELF46F11
c 33 52.8 15.4 192929 60 AC005505
c 34 52.2 15.3 16920 14 SPBC31E1
c 35 52.2 15.3 17855 12 AC004698
c 36 49.2 14.4 65675 71 AC026985
c 37 43.4 12.7 932 53 CNS074H8
c 38 40.8 11.9 3802 92 HSGEBCNA
c 39 38.4 11.2 1048 33 CNS076A0
c 40 37.4 10.9 143172 70 AC026462
c 41 37 10.8 525 4 ACUBIFUS
c 42 36.8 10.8 1187 14 SCFPOLY
c 43 36.6 10.7 173154 77 AC079933
c 44 36.4 10.6 1381 15 ZMU29158
c 45 35.8 10.5 3531 14 SCU44026

```

## ALIGNMENTS

```

RESULT 1
AF318186 AF318186 368 bp mRNA ROD 27-NOV-2000
LOCUS Psammomys obesus beacon mRNA, complete cds.
ACCESSION AF318186
VERSION AF318186.1 GI:11345429
KEYWORDS fat sand rat.
SOURCE Psammomys obesus
ORGANISM Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Gerbillinae;
Psammomys.
REFERENCE 1 (bases 1 to 368)
AUTHORS Collier,G.R., McMillan,J.S., Windmill,K., Walder,K.,
Tenne-Brown,J., de Silva,A., Trevasakis,J., Jones,S., Morton,G.J.,
Lee,S., Augert,G., Civitarese,A. and Zimmet,P.Z.
TITLE A novel gene involved in the regulation of energy balance
JOURNAL Diabetes 49 (11), 1766-1771 (2000)
MEDLINE 20527879
PUBMED 11078442
REFERENCE 2 (bases 1 to 368)
AUTHORS Collier,G.R., McMillan,J.S., Windmill,K., Walder,K.,
Tenne-Brown,J., de Silva,A., Trevasakis,J., Jones,S., Morton,G.J.,
Lee,S., Augert,G., Civitarese,A. and Zimmet,P.Z.
TITLE Direct Submission
JOURNAL Submitted (30-OCT-2000) Metabolic Research Unit, Deakin University,
Pydgons Rd, Geelong, Vic 3217, Australia
FEATURES location/Qualifiers
source 1. 368
/organism="Psammomys obesus"
/db_xref="taxon:48139"
/tissue_type="hypothalamus"
CDS 25..246
/note="involved in control of energy balance"

```

```

/codon_start=1
/product="beacon"
/protein_id="AAG34704.1"
/db_xref="GI:11345430"
/translation="MIEVVCNDRLGKKVRKCNKTDITGDLKLLAAQTGTWRNKIVL
KKWTIFKDHVSLGDIYEIHDGMNLELYQ"
polya_signal 338..343
BASE COUNT 115 a 81 c 78 g 94 t
ORIGIN
Query Match 94.7% Score 324; DB 94; Length 368;
Best Local Similarity 100.0%; Pred. No. 6e-95; 0; Indels 0; Gaps 0;
Matches 324; Conservative 0; Mismatches 0;
Qy 5 caggagattacagctccagcccaaatgattgaggtgtgttgcacacgcgtctaggaag 64
|||||
Db 1 CAGGAGATTACAGCTCCAGCCCAATGATTGAGTGTGTTCGAACCCGCTCTAGGAAG 60
|||||
Qy 65 aaagtcgcgttaagtgcacacccgatgcacacatcggtggaacttgaaagactgatacg 124
|||||
Db 61 AAAGTCGCGGTTAAGTGCACACCGATGACACATCGGGACTTGAAGAACTGATAGCG 120
|||||
Qy 125 gcccaactggcactcgttggaataagatcgcttcttaaaagtggtacacgattttaag 184
|||||
Db 121 GCCCAACTGGCACCCTCGTTGGAAATAGATCGTTCCTTAAAAAGTGTACACGATTTTAAAG 180
|||||
Qy 185 gaccatgatctctctgggagattatgaatccacgatgggatgaacacctgagctttattac 244
|||||
Db 181 GACCATGATCTCTGGGAGATTATGAATCCACGATGGGATGAACCTGGAGCTTTATTAC 240
|||||
Qy 245 cagtagaggggaattctctccacactgcccacactgtcttctctccatggctcat 304
|||||
Db 241 CAGTAGAGGGGAATCTCTCCACTTGCCTTCTCTCTCCATCCCATTCATTTA 300
|||||
Qy 305 acactgtgtagatgactcatttt 328
|||||
Db 301 ACACGTGTTGATAGTCTCATTTT 324

```

```

RESULT 2
AC079644 204394 bp DNA ROD 01-JAN-2001
LOCUS Mus musculus 12 BAC RP23-17406 (Roswell Park Cancer Institute Mouse
DEFINITION BAC Library) complete sequence.
ACCESSION AC079644
VERSION AC079644.9 GI:12000470
KEYWORDS HTG.
SOURCE house mouse.
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE 1 (bases 1 to 204394)
AUTHORS Metzker,M.L., Lewis,J.R., Hume,J., Edwards,C., Harris,C.,
Dederich,D., Thomas,S., Okwuonu,G., Carlock,C., Garner,T.,
Addison,S., Pace,A., Williams,G., Bonnin,D., Brooks,A., Brown,J.,
Buhay,C., Bunac,C., Burkett,C., Chacko,J., Chen,G., Chen,Z.,
Cox,C., Davis,C., Delgado,O., Ding,Y., Dugan-Rocha,S.,
Fernandez,C., Ferraguto,D., Forcum-Tansey,J., Gill,R.,
Gorell,J.H., Gunarathne,P., Haller,G., Hernandez,J., Hogues,M.,
Hosak,H., Hou,X., Huber,J., Jackson,L., Jia,Y., Kelly,J., Kelly,S.,
Kovar,C., Liu,J., Liu,W., Louised,H., Lozado,R.J., Martin,R.,
Massey,E., McLeod,M.P., Mei,G., Moore,S., Morgan,M., Morris,S.,
Neal,D., Nelson,A., Nguyen,R., Shen,H., Vasquez,L., Watlington,S.,
Perez,L., Reiter,D., Say,J., Shen,H., Vasquez,L., Watlington,S.,
Williamson,A., Wrensford,G., Zhou,X., Bouck,J., Hodgson,A.,
Muzny,D.M., Rives,M., Scherer,S., Sodergren,E., Weinstock,G.,
Worley,K. and Gibbs,R.
TITLE Direct Submission
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 204394)
AUTHORS Worley,K.C.
TITLE Direct Submission

```

## JOURNAL

Submitted (07-SEP-2000) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA  
 3 (bases 1 to 204394)  
 Worley K.C.  
 Direct Submission

## REFERENCE

## AUTHORS

## TITLE

## JOURNAL

Submitted (01-JAN-2001) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA

## COMMENT

On Jan 1, 2001 this sequence version replaced gi:11991738. Sequencing is completed to a minimum standard of double strand coverage with a minimum of 2 clones and 2 reads with no ambiguities or 2 chemistries with a minimum of 2 clones and 3 reads with no ambiguities. If the sequence quality does not meet this standard, it will be indicated in the annotation.

The repeat regions shown were identified using RepeatMasker by Adrian Smit.

Sequence similarities were identified using Powerblast by Jinghui Zhang.

Exon/Intron boundaries of identified genes were chosen if there were canonical splice junctions that maintained sequence continuity across the splice junctions.

## FEATURES

## Source

1. 204394  
   Location/Qualifiers  
     /organism="Mus musculus"  
     /db\_xref="taxon:10090"  
     /chromosome="12"  
     /clone="RP23-17406"

repeat\_region

/rpt\_family="L1"

repeat\_region

complement(272..912)

/rpt\_family="Lx2"

repeat\_region

1019..1138

/rpt\_family="B1\_MM"

repeat\_region

1243..1336

/rpt\_family="B3"

complement(1506..1575)

/rpt\_family="PB1"

repeat\_region

2079..2206

/rpt\_family="PB1D10"

repeat\_region

2222..2287

/rpt\_family="RSINE1"

repeat\_region

2288..2310

/rpt\_family="(TTAAA)n"

repeat\_region

2411..2613

/rpt\_family="B3"

repeat\_region

3061..3143

/rpt\_family="T-rich"

complement(3144..3284)

repeat\_region

3285..3302

/rpt\_family="B1\_MM"

repeat\_region

3285..3302

/rpt\_family="T-rich"

repeat\_region

3316..3443

/rpt\_family="B1-F"

repeat\_region

3584..3690

/rpt\_family="B1\_MM"

repeat\_region

3691..3735

/rpt\_family="(CA)n"

repeat\_region

3736..3756

/rpt\_family="B1\_MM"

repeat\_region

4008..4163

/rpt\_family="URR1A"

repeat\_region

5394..5823

/rpt\_family="MT-INTERNAL"

repeat\_region

5851..6040

/rpt\_family="B3A"

repeat\_region

6946..7032

/rpt\_family="B1\_MM"

repeat\_region

7475..7769

/rpt\_family="RLTRETN\_MM"

repeat\_region

repeat\_region

7782..7805

/rpt\_family="(CAGA)n"

repeat\_region

complement(8113..8211)

/rpt\_family="ORR1A2"

repeat\_region

8711..8779

/rpt\_family="(TTTTTC)n"

repeat\_region

complement(8839..9415)

/rpt\_family="L1"

repeat\_region

9419..9656

/rpt\_family="Lx2B"

repeat\_region

10068..10256

/rpt\_family="B2"

repeat\_region

10355..10506

/rpt\_family="(TA)n"

repeat\_region

10690..10728

/rpt\_family="(CAAAA)n"

repeat\_region

11537..11693

/rpt\_family="B1\_MM"

misc\_feature

complement(14141..14699)

/note="Region similar to Mm#S355374 AU015316 Mus musculus CDNA:gb-AU015316"

repeat\_region

15017..16339

/rpt\_family="L1F"

repeat\_region

complement(20743..21668)

/rpt\_family="ORR1A-INT"

repeat\_region

complement(21669..21990)

/rpt\_family="ORR1A2"

repeat\_region

complement(22241..22307)

/rpt\_family="B4"

repeat\_region

22609..22758

/rpt\_family="B3"

repeat\_region

complement(22809..22937)

/rpt\_family="RSINE1"

repeat\_region

complement(23153..23297)

/rpt\_family="B1\_MM"

repeat\_region

complement(23359..23413)

/rpt\_family="B2"

repeat\_region

complement(23900..24050)

/rpt\_family="B1\_MM"

repeat\_region

complement(24051..24168)

/rpt\_family="B2"

repeat\_region

complement(24183..24350)

/rpt\_family="B2"

repeat\_region

complement(24491..24571)

/rpt\_family="B1\_MM"

repeat\_region

24607..24765

/rpt\_family="ORR1A3"

repeat\_region

24770..24876

/rpt\_family="B1\_MM"

repeat\_region

25119..25262

/rpt\_family="(TTCC)n"

repeat\_region

25306..25427

/rpt\_family="MTD"

repeat\_region

25741..25813

/rpt\_family="B3"

repeat\_region

25814..25962

/rpt\_family="Lx7"

repeat\_region

complement(26229..26333)

/rpt\_family="B2"

repeat\_region

complement(26553..26756)

/rpt\_family="B2"

repeat\_region

26842..27153

/rpt\_family="Lx2"

misc\_feature

27339..27790

/note="Region similar to Mm#S189632 vd29g11.s1 Mus musculus CDNA:gb-AA422861"

gene

complement(join(27558..27790,34041..34255,65878..66313,72368..72803,81034..81465,137727..138158,193391..193825))

/gene="g112580559/gb1AF0009871HSAF000987 Homo sapiens eIF-1A, Y isoform (EIF1AY) mRNA"

repeat\_region

complement(35572..35704)

/rpt\_family="B1\_MM"

repeat\_region

complement(35705..35802)

```
repeat_region /rpt_family="B2" complement(35806..35931)
repeat_region /rpt_family="Lx2" complement(35930..36646)
repeat_region /rpt_family="Lx2" complement(36961..37371)
repeat_region /rpt_family="MULV" complement(39203..39223)
repeat_region /rpt_family="RMER2" complement(39224..39617)
repeat_region /rpt_family="MTA" complement(39618..39728)
repeat_region /rpt_family="RMER2" complement(39728..40572)
repeat_region /rpt_family="ORR1A-INT" complement(40582..40880)
repeat_region /rpt_family="ORR1A2" complement(41260..41315)
repeat_region /rpt_family="ORR1B" complement(41320..41920)
repeat_region /rpt_family="L1_MM" complement(41921..42123)
repeat_region /rpt_family="L1"

Query Match 68.6%; Score 234.6; DB 94; Length 204394;
Best Local Similarity 85.7%; Pred. No. 3.2e-65;
Matches 287; Conservative 0; Mismatches 39; Indels 9; Gaps 2;

Qy 2 ttccaggagattacagctccagccacaaatgattgaggtggtttgcaacgacogctctagga 61
Db 49786 TTCAGGAGATTGGAGCTCCAGCCACAATGATTGAGGTGTTTGGCAATGACCGTCTAGGA 49727

Qy 62 aagaaagtcgcgttaagtgaacacccagatgacacacatcggggacctgaagaaactgata 121
Db 49726 AAGAAAGTCCGAGTGAAGTGAACACCCAGATGATACCATTTGGTGTGAAGAAACTGATA 49667

Qy 122 gcgcccacaaatgcacgttgtaagaatgaatcggtcttcttaaaagtgtacacgatttt 181
Db 49666 CGCGCTCAAACTGGCAGCCCGCTTGAATGAAGATCATCTTAAAGATGTTACACGATTAT 49607

Qy 182 aaggaccatgtatctctgggagattatgaatccacgattggaatgaacgtgagctttat 241
Db 49606 AAGGACCACGTGTCTCTGGGAGACTATGAATCCATGATGGATGAACCTGGAGCTTTAT 49547

Qy 242 taccagtagaggggaattctctccacct-----tgcccacactgtttctctctccca 293
Db 49546 TACCAGTAGAGGGAATCTCTCTCTCTGCCCCACTCTACCCCTGCTTCTCTATCCCA 49487

Qy 294 tggctcatttaacactgtgttagatgctcattttt 328
Db 49486 -CCCTCAATTGAAACTTCTATAGATGGTCAATTTT 49453

RESULT 3
AC079644 204394 bp DNA ROD 01-JAN-2001
LOCUS Mus musculus 12 BAC RP23-17406 (Roswell Park Cancer Institute Mouse
DEFINITION BAC Library) complete sequence.
AC079644
AC079644.9 GI:12000470
VERSION
KEYWORDS
SOURCE house mouse.
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 204394)
REFERENCE Metzker, M.L., Lewis, L.R., Hume, J., Edwards, C., Harris, C.,
Dederich, D., Thomas, S., Okwuonu, G., Carlock, C., Garner, F.,
Addison, S., Pace, A., Williams, G., Bonnin, D., Brooks, A., Brown, J.,
Buhay, C., Bunac, C., Burkett, C., Chacko, J., Chen, G., Chen, Z.,
Cox, C., Davis, C., Delgado, O., Ding, Y., Dugan-Rocha, S.,
Fernandez, C., Ferraguto, D., Forcum-Tansey, J., Gill, R.,
Gorrell, J.H., Gunaratne, P., Haller, G., Hernandez, J., Hogues, M.,
```

Hosak, H., Hou, X., Huber, J., Jackson, L., Jia, X., Kelly, J., Kelly, S.,
 Kovar, C., Liu, J., Liu, W., Lou, S., Lou, S., Lou, S., Lou, S., Lou, S.,
 Massey, E., McLeod, M.P., Mei, G., Moore, S., Morgan, M., Morris, R.,
 Neal, D., Nelson, A., Nguyen, R., Nguyen, N., Ogih, M., Parish, B.,
 Perez, L., Reiter, D., Say, J., Shen, H., Vasquez, L., Watlington, S.,
 Williamson, A., Wrensford, G., Zhou, X., Bouck, J., Hodgson, A.,
 Muzny, D.M., Rives, M., Scherer, S., Sodergren, E., Weinstock, G.,
 Worley, K. and Gibbs, R.

Direct Submission  
 Unpublished  
 2 (bases 1 to 204394)  
 Worley, K.C.  
 Direct Submission  
 Submitted (07-SEP-2000) Human Genome Sequencing Center, Department  
 of Molecular and Human Genetics, Baylor College of Medicine, One  
 Baylor Plaza, Houston, TX 77030, USA  
 3 (bases 1 to 204394)  
 Worley, K.C.  
 Direct Submission  
 Submitted (01-JAN-2001) Human Genome Sequencing Center, Department  
 of Molecular and Human Genetics, Baylor College of Medicine, One  
 Baylor Plaza, Houston, TX 77030, USA  
 On Jan 1, 2001 this sequence version replaced gi:11991738.  
 Sequencing is completed to a minimum standard of double strand  
 coverage with a minimum of 2 clones and 2 reads with no ambiguities  
 or 2 chemistries with a minimum of 2 clones and 3 reads with no  
 ambiguities. If the sequence quality does not meet this standard,  
 it will be indicated in the annotation.

The repeat regions shown were identified using RepeatMasker by  
 Adrian Smit.

Sequence similarities were identified using Powerblast by Jinghui  
 Zhang.

Exon/Intron boundaries of identified genes were chosen if there  
 were canonical splice junctions that maintained sequence continuity  
 across the splice junctions.

Location/Qualifiers  
 1. 204394  
 /organism="Mus musculus"  
 /db\_xref="taxon:10090"  
 /chromosome="12"  
 /clone="RP23-17406"  
 2. 273  
 /rpt\_family="L1"  
 complement(272..912)  
 /rpt\_family="Lx2"  
 1019..1138  
 /rpt\_family="B1\_MM"  
 1243..1336  
 /rpt\_family="B3"  
 complement(1506..1575)  
 /rpt\_family="PB1"  
 2079..2206  
 /rpt\_family="PB1D10"  
 2222..2287  
 /rpt\_family="RSINE1"  
 2288..2310  
 /rpt\_family="(TTAA)n"  
 2411..2613  
 /rpt\_family="B3"  
 3061..3143  
 /rpt\_family="T-rich"  
 complement(3144..3284)  
 /rpt\_family="B1\_MM"  
 3285..3302  
 /rpt\_family="T-rich"  
 3316..3443  
 /rpt\_family="B1-F"  
 3584..3690  
 /rpt\_family="B1\_MM"  
 3691..3735



```

repeat_region /rpt_family="(CA)n"
3736..3756
repeat_region /rpt_family="B1_MM"
4008..4163
repeat_region /rpt_family="URR1A"
5394..5823
repeat_region /rpt_family="MT-INTERNAL"
5851..6040
repeat_region /rpt_family="B3A"
6946..7032
repeat_region /rpt_family="B1_MM"
7475..7769
repeat_region /rpt_family="RLTRETIN_MM"
7782..7805
repeat_region /rpt_family="(CAGA)n"
complement(8113..8211)
repeat_region /rpt_family="ORR1A2"
8711..8779
repeat_region /rpt_family="(TTTTTG)n"
complement(8839..9415)
repeat_region /rpt_family="L1"
9419..9656
repeat_region /rpt_family="Lx2B"
10068..10256
repeat_region /rpt_family="B2"
10355..10506
repeat_region /rpt_family="(TA)n"
10690..10728
repeat_region /rpt_family="(CAAAA)n"
11537..11693
repeat_region /rpt_family="B1_MM"
complement(14141..14699)
misc_feature /note="Region similar to Mm#S355374 AU015316 Mus musculus
cDNA:gb-AU015316"
15017..16339
repeat_region /rpt_family="L1F"
complement(20743..21668)
repeat_region /rpt_family="ORR1A-INT"
complement(21669..21990)
repeat_region /rpt_family="ORR1A2"
complement(22241..22307)
repeat_region /rpt_family="B4"
22609..22758
repeat_region /rpt_family="B3"
complement(22809..22937)
repeat_region /rpt_family="RSINE1"
complement(23133..23297)
repeat_region /rpt_family="B1_MM"
complement(23359..23413)
repeat_region /rpt_family="B2"
complement(23900..24050)
repeat_region /rpt_family="B1_MM"
complement(24051..24168)
repeat_region /rpt_family="B2"
complement(24183..24350)
repeat_region /rpt_family="B2"
complement(24491..24571)
repeat_region /rpt_family="B1_MM"
24607..24765
repeat_region /rpt_family="ORR1A3"
24770..24876
repeat_region /rpt_family="B1_MM"
25119..25262
repeat_region /rpt_family="(TTCC)n"
25306..25427
repeat_region /rpt_family="MTD"
25741..25813
repeat_region /rpt_family="B3"
25814..25962
repeat_region /rpt_family="Lx7"
complement(26229..26333)
repeat_region /rpt_family="B2"
complement(26553..26756)

```

```

/rpt_family="B2"
26842..27153
/rpt_family="Lx2"
27339..27790
/note="Region similar to Mm#S189632 vd29g11.s1 Mus
musculus cDNA:gb-AA422861"
complement(join(27558..27790,34041..34255,65878..66313,
72368..72803,81034..81465,137727..138158,193391..193825))
/gene="gii2580559[gb]AF000987[HS]AF000987 Homo sapiens
eIF-1A, Y isoform (EIF1AY) mRNA"
complement(35572..35704)
/rpt_family="B1_MM"
complement(35705..35802)
/rpt_family="B2"
complement(35806..35931)
/rpt_family="Lx2"
complement(35930..36646)
/rpt_family="Lx2"
complement(36961..37371)
/rpt_family="MULV"
39203..39223
/rpt_family="RMER2"
39224..39617
/rpt_family="MTA"
39618..39728
/rpt_family="RMER2"
39728..40572
/rpt_family="ORR1A-INT"
40582..40880
/rpt_family="ORR1A2"
complement(41260..41315)
/rpt_family="ORR1B"
complement(41320..41920)
/rpt_family="L1_MM"
complement(41921..42123)
/rpt_family="L1"

Query Match 68.0%; Score 232.4; DB 94; Length 204394;
Best Local Similarity 85.1%; Pred. No. 1,7e-64;
Matches 286; Conservative 0; Mismatches 41; Indels 9; Gaps 2;

QY 1 gttccaggagattacagctccagccacaaatgattgagtggttgcgaacgacctagg 60
Db 112084 GTTCAGGACATGGAGCTCCAGCCACAAATCATTTGAGTGTGTTGCAATGACCATCTAGG 112143

QY 61 aaagaaagtccggttaagtgcacacccagatgcacacatcggggaactggaagaaactgat 120
Db 112144 AAAGAAAGTCTGAGTGAAGTGAACACCGATGATACCATTTGGTCACTTGAAGAACTGAT 112203

QY 121 agcggcccaactggcactcgttggaataagatcgttctttaaagaagtgtgacagatttt 180
Db 112204 AGCGCTCAAACTGGCACCCCTGGAATTAAGATCATTTCTTAAAAAGTGTGTACACGATTTA 112263

QY 181 taaggaccatgtatctctgggagattatgaaatccacagatggatggaacctggagcttta 240
Db 112264 TAAGGACCACTGTCTCTGGGAGACTATGAAATCCATCATGATGAGGATGAACCTGGAGCTTTA 112323

QY 241 ttaccagtagaggggaattctctccacct-----tgcccaaccttgccttctctctccc 292
Db 112324 TTACCAGTAGAATGAATTTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 112383

QY 293 atggctcatttaacactgtgtgtagatgctcatttt 328
Db 112384 A-CCCTCATTGACACTTCTATAGATGTCATTTTT 112418

RESULT 4
AC021852
LOCUS AC021852 176560 bp DNA HTG 07-JUL-2000
DEFINITION Homo sapiens chromosome 17 clone RP11-474K4, WORKING DRAFT
SEQUENCE, 30 unordered pieces.
ACCESSION AC021852
VERSION AC021852.4 GI:8568867

```

```

KEYWORDS      HTG: HTGS_PHASE1; HTGS_DRAFT.
SOURCE
ORGANISM      Homo sapiens
REFERENCE     Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
AUTHORS       Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
TITLE         Waterston, R.H.
JOURNAL       The sequence of Homo sapiens clone
REFERENCE     1 (bases 1 to 176560)
AUTHORS       Waterston, R.H.
TITLE         Unpublished
JOURNAL       2 (bases 1 to 176560)
AUTHORS       Waterston, R.H.
TITLE         Direct Submission
JOURNAL       Submitted (20-JAN-2000) Genome Sequencing Center, Washington
              University School of Medicine, 4444 Forest Park Parkway, St. Louis,
              MO 63108, USA
COMMENT       On Jun 16, 2000 this sequence version replaced gi:7263413.

----- Genome Center -----
Center: Washington University Genome Sequencing Center
Center code: WUGSC
Web site: http://genome.wustl.edu/gsc/index.shtml
----- Project Information -----
Center project name: H.NH0474K04
----- Summary Statistics -----
Sequencing vector: M13; 74%
Chemistry: Dye-primer ET; 26%
Chemistry: Dye-terminator Big Dye; 26% of reads
Assembly program: Phrap; version 0.990319
Consensus quality: 161959 bases at least Q40
Consensus quality: 165336 bases at least Q40
Consensus quality: 167883 bases at least Q20
Insert size: 168000; agarose-fp
Insert size: 173660; sum-of-contigs
Quality coverage: 3.87 in Q20 bases; agarose-fp
Quality coverage: 3.82 in Q20 bases; sum-of-contigs
-----
* NOTE: This is a 'working draft' sequence. It currently
* consists of 30 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
*
1 1520: contig of 1520 bp in length
* 1521 1620: gap of unknown length
* 1621 3161: contig of 1541 bp in length
* 3162 3261: gap of unknown length
* 3262 5461: contig of 2200 bp in length
* 5462 5561: gap of unknown length
* 5562 7149: contig of 1588 bp in length
* 7150 7249: gap of unknown length
* 7250 11327: contig of 4078 bp in length
* 11328 11428: gap of unknown length
* 11428 13630: contig of 2203 bp in length
* 13631 13730: gap of unknown length
* 13731 15908: contig of 2178 bp in length
* 15909 16008: gap of unknown length
* 16009 18540: contig of 2532 bp in length
* 18541 18640: gap of unknown length
* 18641 20475: contig of 1835 bp in length
* 20476 20576: gap of unknown length
* 20576 22994: contig of 2419 bp in length
* 22995 23095: gap of unknown length
* 23095 26651: contig of 3557 bp in length
* 26652 26752: gap of unknown length
* 26752 31474: contig of 4722 bp in length
* 31474 31574: gap of unknown length
* 31574 35428: contig of 3855 bp in length
* 35429 35529: gap of unknown length
* 35529 40087: contig of 4559 bp in length
* 40088 40187: gap of unknown length

-----
* 40188 44325: contig of 4138 bp in length
* 44326 44426: gap of unknown length
* 44426 49492: contig of 5067 bp in length
* 49492 49592: gap of unknown length
* 49592 57072: contig of 7480 bp in length
* 57072 57172: gap of unknown length
* 57172 63683: contig of 6511 bp in length
* 63683 63783: gap of unknown length
* 63783 70585: contig of 6802 bp in length
* 70585 70686: gap of unknown length
* 70686 76995: contig of 6310 bp in length
* 76995 82834: contig of 5739 bp in length
* 82834 82934: gap of unknown length
* 82934 89274: contig of 6340 bp in length
* 89274 89374: gap of unknown length
* 89374 97794: contig of 8420 bp in length
* 97794 97894: gap of unknown length
* 97894 107952: contig of 10058 bp in length
* 107952 108053: gap of unknown length
* 108053 117316: contig of 9264 bp in length
* 117316 117417: gap of unknown length
* 117417 128399: contig of 10983 bp in length
* 128399 128499: gap of unknown length
* 128499 139336: contig of 11437 bp in length
* 139336 140037: gap of unknown length
* 140037 149771: contig of 9735 bp in length
* 149771 149872: gap of unknown length
* 149872 163648: contig of 13777 bp in length
* 163648 163748: gap of unknown length
* 163748 176560: contig of 12812 bp in length.
* 176560 176560: contig of 12812 bp in length.

FEATURES
Location/Qualifiers
1..176560
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="17"
/clone="RP11-474K4"
1..1520
/note="assembly_name:Contig6"
1621..3161
/note="assembly_name:Contig7"
3262..5461
/note="assembly_name:Contig8"
5562..7149
/note="assembly_name:Contig9"
7250..11327
/note="assembly_name:Contig10"
11428..13630
/note="assembly_name:Contig11"
13731..15908
/note="assembly_name:Contig12"
16009..18540
/note="assembly_name:Contig13"
18641..20475
/note="assembly_name:Contig14
clone_end:T7
vector_side:right"
20576..22994
/note="assembly_name:Contig15"
23095..26651
/note="assembly_name:Contig16"
26752..31473
/note="assembly_name:Contig17"
31574..35428
/note="assembly_name:Contig18"
35529..40087
/note="assembly_name:Contig19"
40188..44325
/note="assembly_name:Contig20"
44426..49492
/note="assembly_name:Contig21"
49593..57072
/note="assembly_name:Contig22"

```



```

* 17857 17956: gap of 100 bp
* 17957 20816: contig of 2860 bp in length
* 20817 20916: gap of 100 bp
* 20917 23798: contig of 2882 bp in length
* 23799 23898: gap of 100 bp
* 23899 26035: contig of 2137 bp in length
* 26036 26135: gap of 100 bp
* 26136 28954: contig of 2819 bp in length
* 28955 29034: gap of 100 bp
* 29055 33642: contig of 4588 bp in length
* 33643 33742: gap of 100 bp
* 33743 36791: contig of 3049 bp in length
* 36792 36891: gap of 100 bp
* 36892 39359: contig of 2468 bp in length
* 39360 39459: gap of 100 bp
* 39460 44800: contig of 5341 bp in length
* 44801 44900: gap of 100 bp
* 44901 50038: contig of 5138 bp in length
* 50039 50138: gap of 100 bp
* 50139 53537: contig of 3399 bp in length
* 53538 53637: gap of 100 bp
* 53638 59913: contig of 6276 bp in length
* 59914 60013: gap of 100 bp
* 60014 65717: contig of 5704 bp in length
* 65718 65817: gap of 100 bp
* 65818 73361: contig of 7544 bp in length
* 73362 73461: gap of 100 bp
* 73462 80860: contig of 7399 bp in length
* 80861 80960: gap of 100 bp
* 80961 89233: contig of 8273 bp in length
* 89234 89333: gap of 100 bp
* 89334 96487: contig of 7154 bp in length
* 96488 96587: gap of 100 bp
* 96588 105683: contig of 9096 bp in length
* 105684 105783: gap of 100 bp
* 105784 113921: contig of 8138 bp in length
* 113922 114021: gap of 100 bp
* 114022 122911: contig of 8890 bp in length
* 122912 123011: gap of 100 bp
* 123012 134520: contig of 11509 bp in length
* 134521 134620: gap of 100 bp
* 134621 144215: contig of 9595 bp in length
* 144216 144315: gap of 100 bp
* 144316 154607: contig of 10292 bp in length
* 154608 154707: gap of 100 bp
* 154708 169066: contig of 14359 bp in length
* 169067 169166: gap of 100 bp
* 169167 185380: contig of 16214 bp in length.

```

## FEATURES

```

source
1. .185380
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="RP11-55J8"
/clone_lib="RPC1-11 Human Male BAC"
1. .1291
/note="assembly_fragment"
1392. .3276
/note="assembly_fragment"
3377. .4744
/note="assembly_fragment"
4845. .6631
/note="assembly_fragment"
6732. .8626
/note="assembly_fragment"
8727. .10349
/note="assembly_fragment"
10450. .11784
/note="assembly_fragment
clone_end:T7
vector_side:right"
11885. .14034
/note="assembly_fragment"
14135. .15586

```

```

misc_feature
/note="assembly_fragment"
15687. .17856
/note="assembly_fragment"
17957. .20816
/note="assembly_fragment"
20917. .23798
/note="assembly_fragment"
23899. .26035
/note="assembly_fragment"
26136. .28954
29055. .33642
/note="assembly_fragment"
33743. .36791
36892. .39359
/note="assembly_fragment
clone_end:SP6
vector_side:right"
39460. .44800
/note="assembly_fragment"
44901. .50038
/note="assembly_fragment"
50139. .53537
/note="assembly_fragment"
53638. .59913
/note="assembly_fragment"
60014. .65717
/note="assembly_fragment"
65818. .73361
/note="assembly_fragment"
73462. .80860
/note="assembly_fragment"
80961. .89233
/note="assembly_fragment"
89334. .96487
/note="assembly_fragment"
96588. .105683
/note="assembly_fragment"
105784. .113921
/note="assembly_fragment"
114022. .122911
/note="assembly_fragment"
123012. .134520
Query Match 57.1%; Score 195.2; DB 64; Length 185380;
Best Local Similarity 78.7%; Pred No. 2.5e-52;
Matches 258; Conservative 0; Mismatches 68; Indels 2; Gaps 2;
Qy 2 ttccaggagattacagctccagccacaatgattgaggtgtttgcaacgacctgtctagga 61
Db 24462 TGCAAGGCAATTTCGAGCTCCAGCTAGGATGATCGAGGTTGTTGCAACCGCTCTGGG 24521
Qy 62 agaaaagtcgcgttaagtgcacaccgatgacaccatcggtgacctggaagaactgata 121
Db 24522 AAAAAAGTCCACGTTAAATGCAACCGGATGATACCATCGGGACCTTAAAGAAGCTGAT 24581
Qy 122 gcgcgccaaactgcacgtcttgaataagatcgcttttaaaaagtggtacacgatttt 181
Db 24582 GCAGCCTAAACATGGGCACCGCTTGGAAACAAGATGTCCTGAAGAAGTGTACACAATTTT 24641
Qy 182 aaggaccatgtatctctgggagattatgaatccacgatgggatgaacctggagctttat 241
Db 24642 AAGGACCATGTGCTCTGGGGGACATATGAATCCACAATGGGATGAACCTGGAGCTTTAT 24701
Qy 242 taccagttagggggaattctctccacctgcccaacctgtcttctctctcccat-ggctca 300
Db 24702 TATCAAAAGATGAGAA-TCCTCATCTTCTGCTGCCCACTCTCTCTCCCATCCCTCACC 24760
Qy 301 ttttaacactgttgtagatgctcatittt 328
Db 24761 CGGCACACTGGGATAGATGCTTGTGT 24788

```

RESULT 6  
AC026620/c

LOCUS  
DEFINITION

ACCESSION  
VERSION

KEYWORDS  
SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

AC026620 186809 bp DNA HTG 18-NOV-2000  
Homo sapiens chromosome 17 clone RP11-443G13 map 17, WORKING DRAFT  
SEQUENCE, 13 unordered pieces.  
AC026620  
AC026620.3 GI:11225463  
HTG; HTGS\_PHASE1; HTGS\_DRAFT.  
human.  
Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
1 (bases 1 to 186809)  
Birren,B., Linton,L., Nusbaum,C., Lander,E., Abraham,H., Allen,N.,  
Anderson,S., Baldwin,J., Barna,N., Bastien,V., Beda,F.,  
Boguslavsky,L., Boukhgaiter,B., Brown,A., Burkett,G.,  
Campiano,A., Castle,A., Choquel,Y., Colangelo,M., Collins,S.,  
Collamore,A., Cooke,P., Dearellano,K., Dewar,K., Diaz,J.S.,  
Dodgson,S., Domino,M., Doyle,M., Ferreira,P., FitzHugh,W., Gage,D.,  
Galagan,J., Gardyna,S., Ginde,S., Goyette,M., Graham,L.,  
Grand-Pierre,N., Grant,G., Hagos,B., Heaford,A., Horton,L.,  
Howland,J.C., Iliev,I., Johnson,R., Jones,C., Kann,L., Karatas,A.,  
Klein,J., LaRocque,K., Lamazares,R., Landers,T., Lehotzky,J.,  
Levine,R., Lieu,C., Liu,G., Locke,K., Macdonald,P., Marquis,N.,  
McCarthy,M., McEwan,P., McGurk,A., McKernan,K., McPheeters,R.,  
Meldrum,J., Meneus,L., Mihova,T., Miranda,C., Mieng,C., Morrow,J.,  
Murphy,T., Naylor,J., Norman,C.H., O'Connor,T., O'Donnell,P.,  
O'Neill,D., Olivari,T.M., Oliver,J., Peterson,K., Pierre,N.,  
Pisani,C., Pollara,V., Raymond,C., Riley,R., Rogov,P., Rothman,D.,  
Roy,A., Santos,R., Schauer,S., Severy,P., Spencer,B.,  
Stange-Thomann,N., Stojanovic,N., Subramanian,A., Talamas,J.,  
Tesfaye,S., Theodore,J., Tirrell,A., Travers,M., Trigilio,J.,  
Vassiliev,H., Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Ye,W.J.,  
Young,G., Zainoun,J., Zilmer,A. and Zody,M.  
Direct Submission  
Submitted (22-MAR-2000) Whitehead Institute/MIT Center for Genome  
Research, 320 Charles Street, Cambridge, MA 02141, USA  
On Nov 18, 2000 this sequence version replaced gi:7534102.  
All repeats were identified using RepeatMasker:  
Smit, A.F.A. & Green, P. (1996-1997)  
http://ftp.genome.washington.edu/RM/RepeatMasker.html  
----- Genome Center  
Center: Whitehead Institute/ MIT Center for Genome Research  
Center code: WIBR  
Web site: http://www-seq.wi.mit.edu  
Contact: sequence\_submissions@genome.wi.mit.edu  
----- Project Information  
Center project name: L8629  
Center clone name: 443-G13  
----- Summary Statistics  
Sequencing vector: M13; M77815; 42% of reads  
Chemistry: Dye-terminator Big Dye; 100% of reads  
Assembly program: Phrap; version 0.960731  
Consensus quality: 182357 bases at least Q40  
Consensus quality: 184169 bases at least Q30  
Consensus quality: 184860 bases at least Q20  
Insert size: 182000; agarose-fp  
Insert size: 185609; sum-of-contigs  
Quality coverage: 10.1 in Q20 bases; agarose-fp  
Quality coverage: 9.9 in Q20.  
\* NOTE: This is a 'working draft' sequence. It currently  
\* consists of 13 contigs. The true order of the pieces  
\* is not known and their order in this sequence record is  
\* arbitrary. Gaps between the contigs are represented as  
\* runs of N, but the exact sizes of the gaps are unknown.  
\* This record will be updated with the finished sequence  
\* as soon as it is available and the accession number will

TITLE

JOURNAL

COMMENT

\* be preserved.  
\* 1 3134: contig of 3134 bp in length  
\* 3135 3234: gap of 100 bp  
\* 3235 6135: contig of 2901 bp in length  
\* 6136 6235: gap of 100 bp  
\* 6236 10291: contig of 4056 bp in length  
\* 10292 10391: gap of 100 bp  
\* 10392 14688: contig of 4297 bp in length  
\* 14689 14788: gap of 100 bp  
\* 14789 24656: contig of 9868 bp in length  
\* 24657 24756: gap of 100 bp  
\* 24757 78004: contig of 53248 bp in length  
\* 78005 89254: gap of 100 bp  
\* 89255 89354: contig of 11150 bp in length  
\* 89355 101823: contig of 12469 bp in length  
\* 101824 101923: gap of 100 bp  
\* 101924 117510: contig of 15587 bp in length  
\* 117511 117610: gap of 100 bp  
\* 117611 135414: contig of 17804 bp in length  
\* 135415 135514: gap of 100 bp  
\* 135515 151939: contig of 16425 bp in length  
\* 151940 152039: gap of 100 bp  
\* 152040 177533: contig of 25494 bp in length  
\* 177534 177634: gap of 100 bp  
\* 177634 186809: contig of 9176 bp in length.  
FEATURES  
Location/Qualifiers  
source  
1. 186809  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/chromosome="17"  
/map="17"  
/clone="RP11-443G13"  
/clone\_lib="RPC1-11 Human Male BAC"  
misc\_feature  
1. 3134  
/note="assembly\_fragment"  
clone\_end:SP6  
vector\_side:left  
3235. 6135  
/note="assembly\_fragment"  
6236. 10291  
/note="assembly\_fragment"  
10392. 14688  
/note="assembly\_fragment"  
14789. 24656  
/note="assembly\_fragment"  
24757. 78004  
/note="assembly\_fragment"  
78105. 89254  
/note="assembly\_fragment"  
89355. 101823  
/note="assembly\_fragment"  
101924. 117510  
/note="assembly\_fragment"  
117611. 135414  
/note="assembly\_fragment"  
135515. 151939  
/note="assembly\_fragment"  
152040. 177533  
/note="assembly\_fragment"  
177634. 186809  
/note="assembly\_fragment"  
clone\_end:T7  
vector\_side:right  
BASE COUNT 48912 a 41810 c 42730 g 52153 t 1204 others  
ORIGIN

Query Match 57.1%; Score 195.2; DB 70; Length 186809;  
Best Local Similarity 78.7%; Pred. No. 2.5e-52;  
Matches 258; Conservative 0; Mismatches 68; Indels 2; Gaps 2;  
QY 2 ttccaggagattacagctccagccacaaatgattgaggttggtgcaacgacctctaggga 61

Db 88227 TCGAAGGCAATTCGAGCTCCAGCTAGGATGATCGAGGTGTTTGCACGACCGCTGCGG 88168  
Oy 62 aagaagtcgcgttaagtcaacacogatgacacccatcggggacttgaagaaactgata 121  
Db 88167 AAAAAGGTCCACGTTAAATGCAACACGCGATACCATCGGGGACCTTAAAGAGCTGATT 88108  
Oy 122 gcggcccaaacctgcactcggttgaaataagatcgcttttaaaagtgtacacgattttt 181  
Db 88107 GCAGCCTAAACTGGCACCCTGTGAACAAAGATGTCCTGGAAGAAGTGTACACAATTTT 88048  
Oy 182 aagaccatgtatctctgggagattatgaatccacatgagatgagacctgagactttat 241  
Db 88047 AAGACCATGTCTCTGCGGAGCTATGAAATCCACAATGGATGAACCTGAGCTTTAT 87998  
Oy 242 taccagtagaggaatctccacccttgcccaacactgctttctctccat-ggctca 300  
Db 87987 TATCAAAAGATGAGAA-TCCCTCATCTCTCTGCTGCCCACTCTCTCTCCCATCCTCACC 87929  
Oy 301 ttttaacctgttagatgctcatttt 328  
Db 87928 CCGCACACTGGGATAGATGCTGTGTTGT 87901  
RESULT 7  
AC074146/c  
LOCUS AC074146 201602 bp DNA HTG 15-JUL-2000  
DEFINITION Mus musculus clone RP23-107G3, WORKING DRAFT SEQUENCE, 38 unordered  
pieces.  
ACCESSION AC074146  
VERSION AC074146.1 GI:9211256  
KEYWORDS HTG; HTGS\_PHASE1; HTGS\_DRAFT.  
SOURCE house mouse.  
ORGANISM Mus musculus  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.  
1 (bases 1 to 201602)  
DOE Joint Genome Institute.  
Sequencing of Mouse  
Unpublished  
2 (bases 1 to 201602)  
DOE Joint Genome Institute.  
Direct Submission  
Submitted (15-JUL-2000) Production Sequencing Facility, DOE Joint  
Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA  
-----Genome Center  
Center: Joint Genome Institute  
Center Code: JGI  
Web site: <http://www.jgi.doe.gov>  
-----  
Project Information  
Center Project Name: 0  
Center clone name: RPCI-23\_107G3  
-----  
Summary Statistics  
Consensus quality: 134464 bases at least Q40  
Consensus quality: 159739 bases at least Q30  
Consensus quality: 170106 bases at least Q20  
Estimated insert size: 181300; agarose-fp estimation  
Quality coverage: 3.41 in Q20 bases; agarose-fp estimation  
Quality coverage: 3.12 in Q20 bases; sum-of-contigs estimation.  
\* NOTE: This is a 'working draft' sequence. It currently  
\* consists of 38 contigs. The true order of the pieces  
\* is not known and their order in this sequence record is  
\* arbitrary. Gaps between the contigs are represented as  
\* runs of N, but the exact sizes of the gaps are unknown.  
\* This record will be updated with the finished sequence  
\* as soon as it is available and the accession number will  
\* be preserved.  
\* 1 1675: contig of 1675 bp in length  
\* 1676 1775: gap of unknown length  
\* 1776 3291: contig of 1516 bp in length

FEATURES

Location/Qualifiers



```

/organism="Mus musculus"
/db_xref="taxon:10090"
/chromosome="11"
/clone="RP23-168P5"
BASE COUNT 56664 a 53067 c 51159 g 56401 t 1936 others
ORIGIN

Query Match 36.8%; Score 125.8; DB 74; Length 219217;
Best Local Similarity 69.0%; Pred. No. 1.3e-29;
Matches 220; Conservative 0; Mismatches 62; Indels 37; Gaps 2;

QY 9 agattacagctccagccacacatgattgagtggttttcaacagcagcgtctagaaagaagaag 68
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 194178 AGATTAAAGTCGCCACACATGATGAGCTGCTTAACAACATCTTCTTGAAGGAAG 194119

QY 69 tccgcgttaagtcaaacacagcagcagcaccatcgaggactggaacaaactgatacgcc 128
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 194118 TCAGGGTTAATGACACAGATGACACCATCAGAACTTGAAGAACTTGATAGCGGCC 194059

QY 129 aaactggcactcgttgaataagatcgttcttaaaagtgatcacacgatttttaaggacc 188
||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
194058 AA-----TGTCAGAGGAGTTTGAAGGACC 194035

QY 189 atgtatctctgggagattatgaatacacagatggatggaactggagctttattaccagt 248
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 194034 ACCTGCTCTCGGAGATTATCAAAATCCACAGTGGGATGAACC-AGAGCTCTGTGACCAAG 193976

QY 249 agagggaattccctccactgcccacactgcttcttcccatggctcattataaac 308
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 193975 AAAGGAGATTCCTCTCTCCCGCCAGCCCTGCTTCTCTCCAGGTTCAATTGACACT 193916

QY 309 tgttgtagtgcctattt 327
||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 193915 GGTAAATAGATTCCTCATTTT 193897

RESULT 9
AC026142/c
LOCUS AC026142 156342 bp DNA HTG 01-JUN-2000
DEFINITION Homo sapiens chromosome 3 clone RP11-601G18 map 3, WORKING DRAFT
SEQUENCE, 13 unordered pieces.
AC026142
VERSION AC026142.2 GI:8139197
KEYWORDS HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
Birren,B., Linton,L., Nusbaum,C. and Lander,E.
Homo sapiens chromosome 3, clone RP11-601G18
Unpublished
2 (bases 1 to 156342)
Birren,B., Linton,L., Nusbaum,C., Lander,E., Abraham,H., Allen,N.,
Anderson,S., Baldwin,J., Barna,N., Bastien,V., Bada,F.,
Boguslavsky,L., Boukhgalter,B., Brown,A., Burkett,G.,
Campopiano,A., Castle,A., Choepel,Y., Colangelo,M., Collins,S.,
Collamore,A., Cooke,P., DeArellano,K., Dewar,K., Diaz,J.S.,
Dodge,S., Domino,M., Doyle,M., Ferreira,P., Fitzhugh,W., Gage,D.,
Galagan,J., Gardyna,S., Glnde,S., Goyette,M., Graham,L.,
Grand-Pierre,N., Grant,G., Hagos,B., Heaford,A., Horton,L.,
Howland,J.C., Iliev,I., Johnson,R., Jones,C., Kann,L., Karatas,A.,
Klein,J., LaRoque,K., Lamazares,R., Landers,T., Lehoczy,J.,
Levine,R., Lieu,C., Liu,G., Locke,K., Macdonald,P., Marquis,N.,
McCarthy,M., McEwan,P., McGurk,A., McKernan,K., McPheeters,R.,
Meidrum,J., Meneus,L., Mihova,T., Miranda,C., Mlenga,V., Morrow,J.,
Murphy,T., Naylor,J., Norman,C.H., O'Connor,T., O'Donnell,P.,
O'Neill,D., Oliver,T.M., Oliver,J., Peterson,K., Pierre,N.,
Pisani,C., Pollara,V., Raymond,C., Riley,R., Rogov,P., Rothman,D.,
Roy,A., Santos,R., Schauer,S., Severy,P., Spencer,B.,
Stange-Thomann,N., Stojanovic,N., Subramanian,A., Talamas,J.,
Tesiaye,S., Theodore,J., Tirrell,A., Travers,M., Trigilio,J.,

```

```

Vassiliev,H., Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Ye,W.J.,
Young,G., Zainoun,J., Zimmer,A. and Zody,M.
Direct Submission
Submitted (19-MAR-2000) Whitehead Institute/MIT Center for Genome
Research, 320 Charles Street, Cambridge, MA 02141, USA
On Jun 1, 2000 this sequence version replaced gi:7264212.
All repeats were identified using RepeatMasker:
Smit, A.F.A. & Green, P. (1996-1997)
http://ftp.genome.washington.edu/RM/RepeatMasker.html
----- Genome Center
Center: Whitehead Institute/ MIT Center for Genome Research
Center code: WIBR
Web site: http://www-seq.wi.mit.edu
Contact: sequence_submissions@genome.wi.mit.edu
----- Project Information
Center project name: L8318
Center clone name: 601_G_18
----- Summary Statistics
Sequencing vector: M13; M7815; 100% of reads
Chemistry: Dye-terminator Big Dye; 100% of reads
Assembly program: Phrap; version 0.960731
Consensus quality: 148258 bases at least Q40
Consensus quality: 152213 bases at least Q30
Consensus quality: 153912 bases at least Q20
Insert size: 163000; agarose-fp
Insert size: 155142; sum-of-contigs
Quality coverage: 4.5 in Q20 bases; agarose-fp
Quality coverage: 4.7 in Q20 bases; sum-of-contigs
-----
* NOTE: This is a 'working draft' sequence. It currently
* consists of 13 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
*
* 1 3818: contig of 3818 bp in length
* 3819 3918: gap of 100 bp
* 3919 8368: contig of 4450 bp in length
* 8369 8468: gap of 100 bp
* 8469 14466: contig of 5998 bp in length
* 14467 14566: gap of 100 bp
* 14567 19098: contig of 4532 bp in length
* 19099 19198: gap of 100 bp
* 19199 25900: contig of 6702 bp in length
* 25901 26000: gap of 100 bp
* 26001 34378: contig of 8378 bp in length
* 34379 34478: gap of 100 bp
* 34479 44257: contig of 9779 bp in length
* 44258 44357: gap of 100 bp
* 44358 54869: contig of 10512 bp in length
* 54870 54969: gap of 100 bp
* 54970 64656: contig of 9687 bp in length
* 64657 64756: gap of 100 bp
* 64757 75758: contig of 11002 bp in length
* 75759 75858: gap of 100 bp
* 75859 98633: contig of 22775 bp in length
* 98634 98733: gap of 100 bp
* 98734 126432: contig of 27699 bp in length
* 126433 126532: gap of 100 bp
* 126533 156342: contig of 29810 bp in length.
FEATURES
Location/Qualifiers
1..156342
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="3"
/map="3"
/clone="RP11-601G18"
/clone_lib="RPC1-11 Human Male BAC"
1..3818
/feature="assembly_fragment"
3919..8368
misc_feature
misc_feature

```



```

/note="assembly_fragment"
8469..14466
/note="assembly_fragment"
14567..119098
/note="assembly_fragment"
clone_end:77
vector_side:left
19199..25900
/note="assembly_fragment"
26001..34378
/note="assembly_fragment"
34479..44257
/note="assembly_fragment"
44358..54869
/note="assembly_fragment"
54970..64656
/note="assembly_fragment"
clone_end:SP6
vector_side:right
64757..75758
/note="assembly_fragment"
75859..98633
/note="assembly_fragment"
98734..126432
/note="assembly_fragment"
126533..156342
/note="assembly_fragment"
47675 a 29500 c 30329 g 47633 t 1205 others
ORIGIN

Query Match      35.5%  Score 121.4;  DB 70;  Length 156342;
Best Local Similarity 71.4%;  Pred. No. 3.4e-28;
Matches 192;  Conservative 0;  Mismatches 66;  Indels 11;  Gaps 2;

QY  2  ttccaggagattacagctccagccacaaatgattgaggtgttggcaacaccctctagga 61
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 103294 TTCAAGGTGATTGAGCTTTAGCTAGGATGATGGAGGTTGTTTGCACCAACTCTCTGGGC 103235

QY  62  aagaaagtcgcgttaagtgcaaccgcatgacaccatcgggacttg-----aa 111
    | |||| |||| |||| |||| |||| |||| |||| |||| |||| |||| |||| |||
Db 103234 TTCAG-GTCCATGTTAAATGCAACATAGATGACGCCACCTTGAAGACCGCTAA 103176

QY  112  gaactgatagcggcccaactggcactcgttggaaatagatgttcttaaaagtggta 171
    ||| |||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
103175 GAAGCTGATTGAGCCCAAGCTGTACCCATTGTAAAGATTATCTTGAGAAATGGTA 103116

172  cagcatttttaggaccatgtatctctgggagattatgaatccacgatgggatgaacct 231
    || |||| |||| |||| |||| |||| |||| |||| |||| |||| |||| ||||
Db 103115 CATGATTTTATAGGACCACACGATATCTGGGGACCATGAATCCATGATGGATGGACCT 103056

QY  232  ggagctttattaccagtagagggaattc 260
    |||| |||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 103055 GGAGCTTTATTTTCAATAGATTAGAATCC 103027

RESULT 10
AC068372/c AC068372 160426 bp DNA HTG 07-JUN-2000
LOCUS Homo sapiens chromosome 3 clone RP11-284K18 map 3, WORKING DRAFT
DEFINITION SEQUENCE, 20 unordered pieces.
ACCESSION AC068372
VERSION AC068372.2 GI:8315072
KEYWORDS HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 160426)
AUTHORS Birren,B., Linton,L., Nusbaum,C. and Lander,E.
TITLE Homo sapiens chromosome 3, clone RP11-284K18
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 160426)

```

AUTHORS

Birren,B., Linton,L., Nusbaum,C., Lander,E., Abraham,H., Allen,N., Anderson,S., Baldwin,J., Barna,N., Bastien,V., Beda,F., Boguslavskiy,L., Boukhgalter,B., Brown,A., Burkett,G., Campopiano,A., Castle,A., Choepel,Y., Colangelo,M., Collins,S., Collymore,A., Cooke,P., DeArellano,K., Dewar,K., Diaz,J.S., Dodge,S., Domino,M., Doyle,M., Ferreira,P., Fitzhugh,W., Gage,D., Galagan,J., Gardyna,S., Ginde,S., Goyette,M., Graham,L., Grand-Pierre,N., Grant,G., Hagos,B., Heaford,A., Horton,L., Howland,J.C., Iliev,I., Johnson,R., Jones,C., Kann,L., Karatas,A., Klein,J., LaRoque,K., Lamazares,R., Landers,T., Lehoczy,J., Levine,R., Lieu,C., Liu,G., Locke,K., Macdonald,P., Marquis,N., McCarthy,M., McEwan,P., McGurk,A., McKernan,K., McPheeters,R., Meldrim,J., Meneus,L., Mihova,T., Miranda,C., Mlenga,V., Morrow,J., Murphy,T., Naylor,J., Norman,C.H., O'Connor,T., O'Donnell,P., O'Neil,D., Oliver,T.M., Oliver,J., Peterson,K., Pierre,N., Pisani,C., Pollara,V., Raymond,C., Riley,R., Rogov,P., Rothman,D., Roy,A., Santos,R., Schauer,S., Severy,P., Spencer,B., Stange-thomann,N., Stojanovic,N., Subramanian,A., Talamas,J., Tesfaye,S., Theodore,J., Tirrell,A., Travers,M., Trigilio,J., Vassiliev,H., Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Ye,W.J., Young,G., Zainoun,J., Zimmer,A. and Zody,M.

Direct Submission

Submitted (02-MAY-2000) Whitehead Institute/MIT Center for Genome Research, 320 Charles Street, Cambridge, MA 02141, USA

On Jun 7, 2000 this sequence version replaced gi:7677743.

All repeats were identified using RepeatMasker:

Smit, A.F.A. & Green, P. (1996-1997)

http://ftp.genome.washington.edu/RM/RepeatMasker.html

----- Genome Center

Center: Whitehead Institute/ MIT Center for Genome Research

Center code: WIBR

Web site: http://www-seq.wi.mit.edu

Contact: sequence\_submissions@genome.wi.mit.edu

----- Project Information

Center project name: L10205

Center clone name: 284\_KJ18

----- Summary Statistics

Sequencing vector: M13; M77815; 100% of reads

Chemistry: Dye-terminator Big Dye; 100% of reads

Assembly program: Phrap; version 0.960731

Consensus quality: 149867 bases at least Q40

Consensus quality: 155113 bases at least Q30

Consensus quality: 157154 bases at least Q20

Insert size: 163000; agarose-fp

Insert size: 158526; sum-of-contigs

Quality coverage: 4.1 in Q20 bases; agarose-fp

Quality coverage: 4.2 in Q20 bases; sum-of-contigs

-----

\* NOTE: This is a 'working draft' sequence. It currently

\* consists of 20 contigs. The true order of the pieces

\* is not known and their order in this sequence record is

\* arbitrary. Gaps between the contigs are represented as

\* runs of N, but the exact sizes of the gaps are unknown.

\* This record will be updated with the finished sequence

\* as soon as it is available and the accession number will

\* be preserved.

\* 1 1063: contig of 1063 bp in length

\* 1064 1163: gap of 100 bp

\* 1164 3028: contig of 1865 bp in length

\* 3029 3128: gap of 100 bp

\* 3129 5167: contig of 2039 bp in length

\* 5168 5267: gap of 100 bp

\* 5268 8327: contig of 3060 bp in length

\* 8328 8427: gap of 100 bp

\* 8428 11683: contig of 3256 bp in length

\* 11684 11783: gap of 100 bp

\* 11784 13964: contig of 2181 bp in length

\* 13965 14064: gap of 100 bp

\* 14065 18527: contig of 4463 bp in length

\* 18528 18627: gap of 100 bp

\* 18628 22883: contig of 4256 bp in length

\* 22884 22983: gap of 100 bp

\* 22984 29442: contig of 6459 bp in length

TITLE

JOURNAL

COMMENT

```

* 29443 29542: gap of 100 bp
* 29543 36305: contig of 6763 bp in length
* 36306 36405: gap of 100 bp
* 36406 43375: contig of 6970 bp in length
* 43376 43475: gap of 100 bp
* 43476 49721: contig of 6246 bp in length
* 49722 49821: gap of 100 bp
* 49822 56627: contig of 6806 bp in length
* 56628 56727: gap of 100 bp
* 56728 66817: contig of 10090 bp in length
* 66818 66917: gap of 100 bp
* 66918 78729: contig of 11812 bp in length
* 78730 78829: gap of 100 bp
* 78830 92103: contig of 13274 bp in length
* 92104 92203: gap of 100 bp
* 92204 104569: contig of 12366 bp in length
* 104570 104669: gap of 100 bp
* 104670 118863: contig of 14194 bp in length
* 118864 118963: gap of 100 bp
* 118964 133390: contig of 14427 bp in length
* 13391 133490: gap of 100 bp
* 133491 160426: contig of 26936 bp in length.
Location/Qualifiers
1. .160426
/organism="Homo sapiens"
/db_xref="taxon:9606"
/map="3"
/clone="RP11-284K18"
/clone_lib="RPC1-11 Human Male BAC"
1. .1063
/note="assembly_fragment"
1184. .3028
/note="assembly_fragment"
3129. .5167
/note="assembly_fragment"
5268. .8327
/note="assembly_fragment"
8428. .11683
/note="assembly_fragment"
11784. .13964
/note="assembly_fragment"
clone_end:SP6
vector_side:right
14065. .18527
/note="assembly_fragment"
clone_end:T7
vector_side:left
18628. .22883
/note="assembly_fragment"
22984. .29442
/note="assembly_fragment"
29543. .36305
/note="assembly_fragment"
36406. .43375
/note="assembly_fragment"
43476. .49721
/note="assembly_fragment"
49822. .56627
/note="assembly_fragment"
56728. .66817
/note="assembly_fragment"
66918. .78729
/note="assembly_fragment"
78830. .92103
/note="assembly_fragment"
92204. .104569
/note="assembly_fragment"
104670. .118863
/note="assembly_fragment"
118964. .133390
/note="assembly_fragment"
133491. .160426

```

```

BASE COUNT 48454 a 31275 c 30247 g 48545 t 1905 others
ORIGIN
Query Match 35.5%; Score 121.4; DB 74; Length 160426;
Best Local Similarity 71.4%; Pred. No. 3.4e-28;
Matches 192; Conservative 0; Mismatches 66; Indels 11; Gaps 2;
Qy 2 ttccaggagattacagctccagcccaaatgattgagtggtttgcaacgacctctcaggga 61
Db 70075 TTCAAGGTGATTGAGCTTTAGCTAGGATGATGGAGGTGTTTGCACACACACGTCCTGGGC 70016
Qy 62 aagaaagtcgcgttaagtgcacaccgatgacaccatcggggacttg-----aa 111
Db 70015 TTCAG-GTCCATGTTAAATGCAACATAGATGACGCCACTGGAGACCTTAAGAGACCGCTAA 69957
Qy 112 gaaactgatacgcccaactggcactcgttggaataagatcgttcttaaaagtgga 171
Db 69956 GAAGCTGATTGCAGCCCAAGCTGTCAACCATTTGTAAGAAATATCTCTGAAGAAATGGTA 69897
Qy 172 cagcatttttaaggaccatgtatctctggagattatgaatccacgattggaacct 231
Db 69896 CATGATTTTGGGACCACACGATGATCTGGGGACCATGAATCCATGATGGGATGGACCT 69837
Qy 232 ggaagtttattaccagtagaggggaattc 260
Db 69836 GGAGCTTTATTTCATAGATTAGAAATCC 69808

RESULT 11
AB023032/c
LOCUS AB023032 59762 bp DNA PLN 27-DEC-2000
DEFINITION Arabidopsis thaliana genomic DNA, chromosome 5, TAC clone:K5J14.
ACCESSION AB023032 BA000015
VERSION AB023032.1 GI:4220631
KEYWORDS
SOURCE Arabidopsis thaliana (strain:Columbia) DNA, clone_lib:Mitsui TAC
clone:K5J14.
ORGANISM
Arabidopsis thaliana
Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;
Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids II;
Brassicales; Brassicaceae; Arabidopsis.
REFERENCE
1 (sites)
AUTHORS Sato,S., Nakamura,Y., Kaneko,T., Kato,T., Asamizu,E., Kotani,H.
and Tabata,S.
TITLE Structural analysis of Arabidopsis thaliana chromosome 5. X.
Sequence features of the regions of 3,076,755 bp covered by sixty
PI and TAC clones
JOURNAL DNA Res. 7 (1), 31-63 (2000)
MEDLINE 20181125
REFERENCE 2 (bases 1 to 59762)
AUTHORS Nakamura,Y.
TITLE Direct Submission
JOURNAL Submitted (01-FEB-1999) Yasukazu Nakamura, Kazusa DNA Research
Institute, Department of Plant Gene Research; 1532-3, Yana,
Kisarazu, Chiba 292-0812, Japan (E-mail:ynakamu@kazusa.or.jp,
Tel:81-438-52-3935, Fax:81-438-52-3934)
COMMENT
Address for correspondence: kaosekazusa.or.jp
For the latest information on annotation of this clone, please see
http://www.kazusa.or.jp/kaos/cgi-bin/agd_graph.cgi?c=K5J14
Genes with similarity to proteins in the databases are described in
'product' or 'note' qualifiers. Genes that have no significant
protein similarity are described as 'unknown protein'.
The software programs used to predict genes include: Grail
(Informatics Group, Oak Ridge National Laboratory,
http://compbio.ornl.gov/Grail-1.3/),
GENSCAN (Chris Burge, MIT, http://CCR-081.mit.edu/GENSCAN.html),
NetGene2 (S.M. Hebsgaard, et al., CBS, Technical University of
Denmark, http://www.cbs.dtu.dk/services/NetGene2/) and
SplicePredictor (Volker Brendel, Stanford University,
http://grenlini.zool.iastate.edu/cgi-bin/sp.cgi).
Genes encoding tRNAs are predicted by tRNAscan-SE

```

(Sean Eddy, Washington University School of Medicine, St. Louis, <http://genome.wustl.edu/eddy/tnAscan-SE/>). This sequence may not be the entire insert of this clone. It may be shorter because we remove overlaps between neighboring submissions. The 5' clone is MJC20 and the 3' clone is MDH9.

# FEATURES

## Location/Qualifiers

```
1..59762
/organism="Arabidopsis thaliana"
/strain="Columbia"
/db_xref="taxon:3702"
/chromosome="5"
/clone="K5J14"
/clone_lib="Mitsui TAC"
join(6320)..6469,6562..6764,6945..7020,7279..7365,
7444..7542)
/notes="contains similarity to tetracycline transporter
protein"
```

## CDS

```
gene_id:K5J14.1"
/codon_start=1
/evidence-not_experimental
/protein_id="BAB10194.1"
/db_xref="GI:10177006"
/translation="MKVFSADQADNVPEGRASAFGLTGTSCAFVCAVLSARELS
IGVTFQVRAAMGILSTLWRLFLPDSIRNSLGAPVINEKLSPLLEDCPGRNRIIF
RAIRLVHMASLRSSVPLQVAMVSFFSLAEAGHASSMYLLKAKFHFKDQFADL
MTIVGASGISQLPMPVLPVLPALKERLLSLGFFGCAHRTVF"
join(11437)..11634,11729..11846,12151..12249,12342..12457,
12552..12638,12814..13008,13096..13165,13238..13392,
13524..13688,13766..13837,13909..14026,14104..14346,
14420..14568,14652..14843,14965..15495,15720..15857,
15937..15999)
/notes="gene_id:K5J14.2
pir||T30561
similar to unknown protein"
```

## CDS

```
/codon_start=1
/evidence-not_experimental
/protein_id="BAB10195.1"
/db_xref="GI:10177007"
/translation="MRSPFCNMALQALCNFLNLTGFMEDQPINOCSSSTNASEKTPPE
STLEHNTLDSRTYFOVKNETVLLFKEKTASTGVPGVGGOORLIFRGRVLKDDHPL
SEYHLENGTLLHIVROAESAPSSGTSGQATDNNTNGSPSRGHVSHSVYLG
SRNGDQTEGIVPDLRSVIGVNSFGVSGQLPTNHSTNGTOSPSNNSNAPPGT
SDGPEIQSQATGHSOPQAPPGVFSQTSMPRVQIPVTAATIPISFTLPIPDS
LDLMEFTNMEQALISQNGYOPTDSAGSGRPRELPRNRGAATPEALSIVLRNAQ
HLLSGPSSLSHIAAGLEODGSSDPTLRSOIOTEAQOVGLAMOHGLAHLLELGTI
LTLRAPSPELSYVAGPAAVYISPSGPNPIMVQRPHOISPLFTGATVSSNPLTGPVG
LCTAORHINIHAGTSGSPMLSSVGNQNSGEGGQDRDSTNSVPAAPVPSHSTGEN
VSAGVQGLDDVSVQAQINARIMVNIHQGRDQIPSGIESLERDMSTGHGVATAMPE
QPTNATTCAPESSGSLHDLPSERSNSVCNEKDLGDLGHEPARAKDTSCTGQSSA
PSGDATGAKETNKATPEVATATPLGLGLDRKRKSKPKVSGKTESDGTSAITLEG
VQSSGTSQLQLSLFSGSSRSDFTLRGOCSDRDVDSAMSQVLESPLVDGLLA
GVSRQGVDSNMLRMQLQFTQNPQIMNTVOOIAQVQDGEIENMSGGAQEGGGF
DPSRVQOMPLVSRFSGGQLPPLPIQPDPRPSQVNVQSMQAIENHSDPPEDVF
RAMVENAISQDELVELNCCDALSOEYAEALLRDLGRLKDDQL"
join(20293)..20530,20628..20735,20827..20919,21008..21100,
21179..21264,21337..21615,21684..21817,21912..22004,
22093..22210,22294..22461)
/notes="gene_id:K5J14.3"
/codon_start=1
/evidence-not_experimental
/product="serine carboxypeptidase-II like"
/protein_id="BAB10196.1"
/db_xref="GI:10177008"
/translation="WAIYSLRDVAMVMTVQVFARGYPETDLVLRPGQPKVFRQYA
GVLDLDMNSRLFTYFVEAEKHPDTPKLTLLWNGPGGSSVGGGAFTBELFPYPTGY
GGLRLNMSGNKSNLLFVDSVAGVGSYSNRSYDNAGDSADMLVFLTRWFDK
FPELKHDLFTGESAGVYIPOLADAILSYNSRSGFKFNKGAIGNPLKLDLDDI
PAYVEFFSHGMSVWGTIKIQCDFSHYTYAYPHNVSDACNDIAIREAGDITTEVYN
TFDLPDLCLYSIALQELRLKQATKMSGVDCMYERQFYLNIPEVOMALHARTN
LPSYSLSNLLNYSADVNTNMLFTLKRIONKIPRIFSGQDSVYFPPLGRTIVG
ELANDLNEKTTVPYGVPHKQVGGWAIEYGNLLTFATVGAHAHAYVTPQGRALHLF
STFLRQRLPNKNTIAMHD"
```

## CDS

```
join(27261)..27510,27591..27698,27785..27877,27952..28044,
28143..28228,28314..28592,28670..28803,28922..29014,
29085..29202,29308..29475)
/notes="gene_id:K5J14.4"
/codon_start=1
/evidence-not_experimental
/product="serine carboxypeptidase II-like"
/protein_id="BAB10197.1"
/db_xref="GI:10177009"
/translation="MASVSRVAVAMVVLVLSQWFAKGYPEEDLVVRLPGOPTVGF
KOYAGYVDVVKAGRSFYVYVEAVKQPKDPLTLWNGPGGSSIGGGAFTELGPFY
PTGDGRLRVNSMSWNKASHLLFVESAGVGSYSNKSQDNTGDKSTANDMLVELLR
WPKRFPKLKSRDLFLTGESYAGHVPOLADAILSYNSHSGFKENIGVAGNELLKL
DRSPATYEEFWSHGMSIDELKLTITISQCDFTDYTFASPHNVSNACNEASENIT
EYNNYDVLVDCGVYSYVQELRLKMATKMSGVDCMYERFYNLPEVQKALHA
NRTHLPYSWMSQSVLNSDIDGNIDMLPLIKRLILNKTPWIFSGQDSVVPFGGSR
TLVRELAQDLNFKTTPVYGAWFHKSQVGGWAIEYGLKLTATVGAHMYVPAQPSRA
LHLFSSPVSGRRLPNTHSSDTE"
join(32631)..32691,32788..32924,33757..33803,34429..34757,
34840..34922,35182..35257,35373..35437,35509..35604,
35716..35994)
/notes="gene_id:K5J14.5"
/codon_start=1
/evidence-not_experimental
/product="alcohol dehydrogenase"
/protein_id="BAB10198.1"
/db_xref="GI:10177010"
/translation="MENGNSSSDNKSHKPLRCKAAVSRKAGEPLVMEIIVAPPQPF
EVRIRICTALCHSDVTFMKLOVPPACFPRLTGLHAGVSVGCVENKVVVGDTVLP
TFMPDCDVCDSKSHKNSCLSGPPFKVSPMRYDNNRSRTDLNGETLFLFLNVSFS
EYTVLDVAVNKKIDSSIPRSCALLSCSGVSGVGAWETAKVKGSTVWIFGLSGIGL
AVAEGARLGCASRIIGVDINPTFKQGVKGFVTFVNSMTCERNKRVSEVINEMDGA
DYCECVGSSSLVQEAACQCGMGKTITLGVDPKSGICLDSFDVLHKGILMGSIF
GLKAKTHIPILKRLYLSNELELDKFTVHEMKPEINDAFQLLLEGKRCIRCLVWG"
37256..38779
/notes="gene_id:K5J14.6"
/codon_start=1
/evidence-not_experimental
/product="beta-glucosidase"
/protein_id="BAB10199.1"
/db_xref="GI:10177011"
/translation="MRTIYLSLFIIVLALNEVMKSHSTPKLRSDPDPEDIFGA
ATSAVOGAAHEDGRPSIWDTFSEKYPKIDGNSGSIASDSHYLKEDVGLLHQI
GFAYRFSISWRILPRENLKGGINAGIDYNNLILNELSKGTPFATPHMDTPQS
LEDAYGFLGAEIVNDFRADICFKNFGVRHMTLNEPLTVQOQYVYVAGMAPGR
CSFTNPNCAGNATGATPYVGHNLILAHGEAVKVRKYKASOKGVGTALNAGWNL
PYSSEADRLAAARAMAFEDYFMEPLVTKYPIDMVNVYVKGRLPTFTAKQSKMLKG
SYDFIGNYTSSTAKDVPCCSENVLISDPCASVTGERGVPVIGRAASDWLLIYPK
GIRDLLLAKYKFKDPVYITENGRDEASTGKIDLKDSEIDYTAQHLKMWQDAISIG
ANYKGFATSLLDNFENWATGYAVRGLVYVDFNGRKRYPKSKAKFKLLNERKKN"
join(41203)..42492,42572..42818,42900..43083,43173..43344,
43425..43646)
/notes="gene_id:K5J14.7"
/codon_start=1
/evidence-not_experimental
/product="cell division protein FtsH"
/protein_id="BAB10200.1"
/db_xref="GI:10177012"
/translation="MATTSSNPLLLSNFLGSQIIISAPTPKTTTKSLPSVVISRKY
QISQSEKMLKSLPSQAALALFFSSSPQALAVNEPVQPPATITABAQSNLSTFGQ
NVLMTAPNQASQSDLPDGTQWRYSEFLNAVKKGVKRVFSDKGSYQLTAVDNRA
TVIVPNPDLDILAMNGVDISVSEGGNGLFDIGNLLFLPLAFGLFYLRGGOG
GAGGPGGLGPMDFGRSKFOEVPETGVTFGVAGADQAKLEQVYVDFLKNPKDKT
ALGAKIPKGLLVGPPCTGKTLARAVAGAVPFFSCAASEFEVLELVGVGASVRDL
FERAKSKACIVFIDEIDAVGRQAGCGGNDREQTINQLITMEFGSGNSGVIVL
APNRPDVLALLRPGFRQVTPDVPVAGVOILKRVHSRGAIKQDVKDYVARR
TQGTGADLQNLMEARILAAARELKESIDALEISDALRIAGPEKNNAVYSEKRL
MAYEHAGHALVAMPEDYVTKASIIIPRQAGLTTFAPSERLSGLSYRSYLENO
MAYEHAGHAEEVIFGDEYVTPVITGASNDMOVSRVAROMFERFSGKIGOVAGGAG
NPLGSSMSQKDYSNATADVDAEVELVEKAYVRAKEIITTDIDILHLKLAQLLIEK
ETVDGSENFSLFDGOALEYVS"
join(44029)..45531,45787..45882,46066..46551)
/notes="contains similarity to CHP-rich zinc finger
protein"
```

## CDS



```
intron      complement(6937..7033)
exon        /number=3
complement(7034..7426)
intron      /number=4
complement(7427..7614)
exon        /number=4
complement(7615..7696)
intron      /number=5
complement(7697..8279)
exon        /number=5
complement(8280..8578)
intron      /number=6
complement(8579..9066)
exon        /number=6
complement(9067..9216)
LTR         /number=7
14402..14473
/note="72 bp long terminal repeat"
14495..18651
/note="similarity to fragments of copia-like polyprotein"
LTR         19002..19073
/note="72 bp long terminal repeat"
exon        19415..19534
/gene="T14D3.40"
/nnumber=1
19415..21625
/gene="T14D3.40"
CDS         join(19415..19534,19838..20410,20508..20662,20758..20837,
20930..21066,21169..21272,21360..21450,21533..21625)
/gene="T14D3.40"
/note="strong similarity to
n-acetylglucosaminyl-phosphatidylinositol -
Schizosaccharomyces pombe, PIR:"40367"
/product="n-acetylglucosaminyl-phosphatidylinositol-like
protein"
/protein_id="CAB72148.1"
/db_xref="GI:6911848"
/translation="MAEPPKRLVLMVSDPFFNFNGVENHYIYLSQCLLKLGKVVVM
THAYGNRSQVMTGLKVVYVPRPFVMTPTPTVYGLPIVRLILREKTVVVGH
QAFSLCHVALMHAETMGYKVVFTDHSLYGFADVGSIHNNKVLQSLADIDAICVSH
TSKNTILKRGSLSPAKVFMIPNAVDTAMFKPVRLEEMREKHSLODRVEMLGAVPHSVRYSLVT
VEVPEVRLYFNRVYGGDGPKHVRLEEMREKHSLODRVEMLGAVPHSVRYSLVT
GHIFNLSLTAFCIAILEAASGLITVSTRVGGFLHGLQVLPDQVLAEPDQDMV
RAIEKAISLTPTINPEEMHNMKLYSWQDVAKTEIVYDRALKCSNRLRLRMREL
SCGANAGLFCVMWILDYLLWRLLQLQPDDEIEAPDCLCHHRGVEVSEGLRKIK
"
19535..19837
/gene="T14D3.40"
/nnumber=1
19838..20410
/gene="T14D3.40"
/nnumber=2
20411..20507
/gene="T14D3.40"
/nnumber=2
20508..20662
/gene="T14D3.40"
/nnumber=3
20663..20757
/gene="T14D3.40"
/nnumber=3
20758..20837
/gene="T14D3.40"
/nnumber=4
20838..20929
/gene="T14D3.40"
/nnumber=4
20930..21066
/gene="T14D3.40"
/nnumber=5
21067..21168
/gene="T14D3.40"
intron      complement(6937..7033)
exon        /number=3
complement(7034..7426)
intron      /number=4
complement(7427..7614)
exon        /number=4
complement(7615..7696)
intron      /number=5
complement(7697..8279)
exon        /number=5
complement(8280..8578)
intron      /number=6
complement(8579..9066)
exon        /number=6
complement(9067..9216)
LTR         /number=7
14402..14473
/note="72 bp long terminal repeat"
14495..18651
/note="similarity to fragments of copia-like polyprotein"
LTR         19002..19073
/note="72 bp long terminal repeat"
exon        19415..19534
/gene="T14D3.40"
/nnumber=1
19415..21625
/gene="T14D3.40"
CDS         join(19415..19534,19838..20410,20508..20662,20758..20837,
20930..21066,21169..21272,21360..21450,21533..21625)
/gene="T14D3.40"
/note="strong similarity to
n-acetylglucosaminyl-phosphatidylinositol -
Schizosaccharomyces pombe, PIR:"40367"
/product="n-acetylglucosaminyl-phosphatidylinositol-like
protein"
/protein_id="CAB72148.1"
/db_xref="GI:6911848"
/translation="MAEPPKRLVLMVSDPFFNFNGVENHYIYLSQCLLKLGKVVVM
THAYGNRSQVMTGLKVVYVPRPFVMTPTPTVYGLPIVRLILREKTVVVGH
QAFSLCHVALMHAETMGYKVVFTDHSLYGFADVGSIHNNKVLQSLADIDAICVSH
TSKNTILKRGSLSPAKVFMIPNAVDTAMFKPVRLEEMREKHSLODRVEMLGAVPHSVRYSLVT
VEVPEVRLYFNRVYGGDGPKHVRLEEMREKHSLODRVEMLGAVPHSVRYSLVT
GHIFNLSLTAFCIAILEAASGLITVSTRVGGFLHGLQVLPDQVLAEPDQDMV
RAIEKAISLTPTINPEEMHNMKLYSWQDVAKTEIVYDRALKCSNRLRLRMREL
SCGANAGLFCVMWILDYLLWRLLQLQPDDEIEAPDCLCHHRGVEVSEGLRKIK
"
```

```

/number=5
21169..21272
/gene="T14D3.40"
/number=6
21273..21359
/gene="T14D3.40"
/number=6
21360..21450
/gene="T14D3.40"
/number=7
21451..21532
/gene="T14D3.40"
/number=7
21533..21625
/gene="T14D3.40"
/number=8
21937..23299
/gene="T14D3.50"
/number=1
22937..23299
/gene="T14D3.50"
22937..23299
CDS
Query Match      32.5%  Score 111;  DB 13;  Length 88010;
Best Local Similarity 68.6%  Pred. No. 7.7e-25;
Matches 153;  Conservative 0;  Mismatches 70;  Indels 0;  Gaps 0;

QY  29  atattgagtggtttgcaacgacctctaggaagaagtcggttaactgcaacacc 88
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db  53979  ATGATAGAGGTGGTGTAAATGATCGGCTTGGGAAGAGTGAGTGAACGAG 53920

QY  89  gatgacaccatcggggacttgaagaactgatagcgcacaaactggcaactcgttggaa 148
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db  53919  GAAGACACCATCGGAGATCTCAAGAAGCTTGTGGGCGCCCAACTGGACCCGCCAA 53860

QY  149  agatgcgttcttaaaaagtggtacacgatttttaagaccatgtatctctgggagattat 208
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db  53859  AAGATTCCGATCCAAAATGGTAGAACATCTACAAGGATCATATCCCTCTCAAGGACTAC 53800

QY  209  gaaatccacgatgggatgaacctggagctttattaccagtaga 251
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db  53799  GAAATCCATGACGCGATGGGTCTCGAGCTCTACTACAACTAAA 53757

RESULT 13
AC016808/c  AC016808  182320 bp  DNA  HTG  22-JUL-2000
LOCUS      Homo sapiens clone RP11-11G12, WORKING DRAFT SEQUENCE, 17 unordered
DEFINITION pieces.
ACCESSION  AC016808
VERSION    AC016808.4  GI:9369471
KEYWORDS   HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE     human.
ORGANISM   Homo sapiens
REFERENCE  1 (bases 1 to 182320)
AUTHORS    Birren,B., Linton,L., Nusbaum,C., Lander,E., Allen,N., Anderson,M.,
           Brown,A., Castle,A., Collins,S., Collins,S., Collins,S., Collins,S.,
           Cooke,P., DeArrellano,K., Dewar,K., Domino,M., Donelan,L., Doyle,M.,
           Ferreira,P., FitzHugh,W., Forrest,C., Funke,R., Gage,D.,
           Galagan,J., Gardyna,S., Grant,G., Hagos,B., Heaford,A., Horton,L.,
           Howland,J.C., Johnson,R., Jones,C., Kann,L., Karatas,A., Klein,J.,
           Lehoczy,J., Lieu,C., Locke,K., Macdonald,P., Marquis,N.,
           McEwan,P., McGurk,A., McKernan,K., McLaughlin,J., Meldrim,J.,
           Morrow,J., Naylor,J., Norman,C.H., O'Connor,T., O'Donnell,P.,
           Peterson,K., Pollara,V., Riley,R., Roy,A., Santos,R., Severy,P.,
           Stange-Thomann,N., Stojanovic,N., Subramanian,A., Talamas,J.,
           Baldwin,J., Barna,N., Beckerly,R., Boguslavsky,L., Boukhgalter,B.,
           Brown,A., Castie,A., Colangelo,M., Collins,S., Collins,S., Collins,S.,
           Cooke,P., DeArrellano,K., Dewar,K., Domino,M., Donelan,L., Doyle,M.,
           Ferreira,P., FitzHugh,W., Forrest,C., Funke,R., Gage,D.,
           Galagan,J., Gardyna,S., Grant,G., Hagos,B., Heaford,A., Horton,L.,
           Howland,J.C., Johnson,R., Jones,C., Kann,L., Karatas,A., Klein,J.,
           Lehoczy,J., Lieu,C., Locke,K., Macdonald,P., Marquis,N.,
           McEwan,P., McGurk,A., McKernan,K., McLaughlin,J., Meldrim,J.,
           Morrow,J., Naylor,J., Norman,C.H., O'Connor,T., O'Donnell,P.,
           Peterson,K., Pollara,V., Riley,R., Roy,A., Santos,R., Severy,P.,
           Stange-Thomann,N., Stojanovic,N., Subramanian,A., Talamas,J.,
```

Tesfaye,S., Tirrell,A., Vassiliev,H., Vo,A., Wheeler,J., Wu,X., Wyman,D., Ye,W.J., Zimmer,A. and Zody,M.

Submitted (07-DEC-1999) Whitehead Institute/MIT Center for Genome Research, 320 Charles Street, Cambridge, MA 02141, USA

On Jul 22, 2000 this sequence version replaced gi:8137597.

All repeats were identified using RepeatMasker:  
Smit, A.F.A. & Green, P. (1996-1997)  
<http://ftp.genome.washington.edu/RM/RepeatMasker.html>

----- Genome Center -----  
Center: Whitehead Institute/ MIT Center for Genome Research  
Center code: WIBR  
Web site: <http://www-seq.wi.mit.edu>  
Contact: [sequence\\_submissions@genome.wi.mit.edu](mailto:sequence_submissions@genome.wi.mit.edu)  
----- Project Information -----  
Center project name: L3153  
Center clone name: 1L\_G\_12  
----- Summary Statistics -----  
Sequencing vector: M13; M7815; 95% of reads  
Sequencing vector: Plasmid; n/a; %-0.f%% of reads  
4.62046204620462Chemistry: Dye-terminator Big Dye; 100% of reads

Assembly program: Phrap; version 0.960731  
Consensus quality: 172181 bases at least Q40  
Consensus quality: 177293 bases at least Q30  
Consensus quality: 179165 bases at least Q20  
Insert size: 185000; agarose-fp  
Insert size: 180720; sum-of-contigs  
Quality coverage: 4.1 in Q20 bases; agarose-fp  
Quality cov.

\* NOTE: This is a 'working draft' sequence. It currently  
\* consists of 17 contigs. The true order of the pieces  
\* is not known and their order in this sequence record is  
\* arbitrary. Gaps between the contigs are represented as  
\* runs of N, but the exact sizes of the gaps are unknown.  
\* This record will be updated with the finished sequence  
\* as soon as it is available and the accession number will  
\* be preserved.

1	1052:	contig of 1052 bp in length
1053	1152:	gap of 100 bp
1153	2627:	contig of 1475 bp in length
2628	2727:	gap of 100 bp
2728	6363:	contig of 3636 bp in length
6364	6463:	gap of 100 bp
6464	10575:	contig of 4112 bp in length
10576	10675:	gap of 100 bp
10676	16048:	contig of 5373 bp in length
16049	16148:	gap of 100 bp
16149	22232:	contig of 6084 bp in length
22233	22332:	gap of 100 bp
22333	28640:	contig of 6308 bp in length
28641	28740:	gap of 100 bp
28741	35908:	contig of 7168 bp in length
35909	36008:	gap of 100 bp
36009	41361:	contig of 5353 bp in length
41362	41461:	gap of 100 bp
41462	48668:	contig of 7207 bp in length
48669	48768:	gap of 100 bp
48769	58942:	contig of 10174 bp in length
58943	59042:	gap of 100 bp
59043	68408:	contig of 9366 bp in length
68409	68508:	gap of 100 bp
68509	81914:	contig of 13406 bp in length
81915	82014:	gap of 100 bp
82015	93085:	contig of 11071 bp in length
93086	93185:	gap of 100 bp
93186	108502:	contig of 15317 bp in length
108503	108602:	gap of 100 bp
108603	129972:	contig of 21370 bp in length
129973	130072:	gap of 100 bp
130073	182320:	contig of 52248 bp in length.

Location/Qualifiers  
1..182320









NUMBER OF SEQ ID NOS: 57  
 SOFTWARE: FastSeq for Windows Version 3.0  
 SEQ ID NO 2  
 LENGTH: 498  
 TYPE: PRT  
 ORGANISM: HIV  
 US-08-889-841-2

Query Match 14.8%; Score 58.5; DB 3; Length 498;  
 Best Local Similarity 36.5%; Pred. No. 7.5;  
 Matches 19; Conservative 7; Mismatches 17; Indels 9; Gaps 2;

QY 20 TDDTIGDLKLLIAAQTGTWN---KIVLKKW-----YTFKDHVSLGDIYEI 62  
 Db 298 TGDGIRDIRQAHCNISGAKWNTLKKVVKLKEQFPNKTIVFNHSSGGDPEI 349

ULT 6  
 US-08-889-841-5  
 Sequence 5, Application US/08889841B  
 Patent No. 6090392  
 GENERAL INFORMATION:  
 APPLICANT: Berman, Phillip W.  
 TITLE OF INVENTION: HIV ENVELOPE POLYPEPTIDES AND VACCINE  
 FILE REFERENCE: 14918-703CIP  
 CURRENT APPLICATION NUMBER: US/08/889,841B  
 CURRENT FILING DATE: 1997-07-08  
 EARLIER APPLICATION NUMBER: US 60/676,737  
 EARLIER FILING DATE: 1996-07-08  
 NUMBER OF SEQ ID NOS: 57  
 SOFTWARE: FastSeq for Windows Version 3.0  
 SEQ ID NO 5  
 LENGTH: 498  
 TYPE: PRT  
 ORGANISM: HIV  
 FEATURE:  
 NAME/KEY: VARIANT  
 LOCATION: (1)---(498)  
 OTHER INFORMATION: Xaa = Any Amino Acid  
 US-08-889-841-5

Query Match 14.8%; Score 58.5; DB 3; Length 498;  
 Best Local Similarity 34.6%; Pred. No. 7.5;  
 Matches 18; Conservative 8; Mismatches 17; Indels 9; Gaps 2;

QY 20 TDDTIGDLKLLIAAQTGTWNK-----IVLKKWY---TIFKDHVSLGDIYEI 62  
 Db 298 TGDGIRDIRQAHCNISGAKWNTLKKVVKLKEQFPNKTIVFNHSSGGDPEI 349

RESULT 7  
 US-08-889-841-31  
 Sequence 31, Application US/08889841B  
 Patent No. 6090392  
 GENERAL INFORMATION:  
 APPLICANT: Berman, Phillip W.  
 TITLE OF INVENTION: HIV ENVELOPE POLYPEPTIDES AND VACCINE  
 FILE REFERENCE: 14918-703CIP  
 CURRENT APPLICATION NUMBER: US/08/889,841B  
 CURRENT FILING DATE: 1997-07-08  
 EARLIER APPLICATION NUMBER: US 60/676,737  
 EARLIER FILING DATE: 1996-07-08  
 NUMBER OF SEQ ID NOS: 57  
 SOFTWARE: FastSeq for Windows Version 3.0  
 SEQ ID NO 31  
 LENGTH: 483  
 TYPE: PRT  
 ORGANISM: HIV  
 US-08-889-841-31

Query Match 14.7%; Score 58; DB 3; Length 483;  
 Best Local Similarity 27.5%; Pred. No. 8.4; 25; Indels 20; Gaps 4;  
 Matches 22; Conservative 13; Mismatches 13; Indels 20; Gaps 4;

QY 2 IEVVCNDRGLGKVKVKN-----TDDTIGDLKLLIAAQTGTWNKIV-----LKK 46  
 Db 255 VKINCT-RLGNNTFKSINIGRVLVYATGEIIGDIRQAHCNISRAQWNTLEKVVVDLKR 313

QY 47 WY-----TIFKDHVSLGDIYEI 62  
 Db 314 QFGDNTTIAFNRSRSGDPEI 333

RESULT 8  
 US-08-889-841-33  
 Sequence 33, Application US/08889841B  
 Patent No. 6090392  
 GENERAL INFORMATION:  
 APPLICANT: Berman, Phillip W.  
 TITLE OF INVENTION: HIV ENVELOPE POLYPEPTIDES AND VACCINE  
 FILE REFERENCE: 14918-703CIP  
 CURRENT APPLICATION NUMBER: US/08/889,841B  
 CURRENT FILING DATE: 1997-07-08  
 EARLIER APPLICATION NUMBER: US 60/676,737  
 EARLIER FILING DATE: 1996-07-08  
 NUMBER OF SEQ ID NOS: 57  
 SOFTWARE: FastSeq for Windows Version 3.0  
 SEQ ID NO 33  
 LENGTH: 487  
 TYPE: PRT  
 ORGANISM: HIV  
 US-08-889-841-33

Query Match 14.7%; Score 58; DB 3; Length 487;  
 Best Local Similarity 27.5%; Pred. No. 8.5;  
 Matches 22; Conservative 13; Mismatches 25; Indels 20; Gaps 4;

QY 2 IEVVCNDRGLGKVKVKN-----TDDTIGDLKLLIAAQTGTWNKIV-----LKK 46  
 Db 255 VKINCT-RLGNNTFKSINIGRVLVYATGEIIGDIRQAHCNISRAQWNTLEKVVVDLKR 313

QY 47 WY-----TIFKDHVSLGDIYEI 62  
 Db 314 QFGDNTTIAFNRSRSGDPEI 333

RESULT 9  
 US-08-022-835-6  
 Sequence 6, Application US/08022835  
 Patent No. 5420030  
 GENERAL INFORMATION:  
 APPLICANT: Reitz Jr., Marvin S.  
 APPLICANT: Franchini, Genoveffa  
 APPLICANT: Markham, Phillip D.  
 APPLICANT: Gallo, Robert C.  
 APPLICANT: Loria, Franco C.  
 APPLICANT: Popovic, Mikulas  
 APPLICANT: Gaerter, Suzanne  
 TITLE OF INVENTION: MOLECULAR CLONES OF HIV-1 AND USES  
 NUMBER OF SEQUENCES: 6  
 CORRESPONDENCE ADDRESS:  
 ADDRESSEE: CUSHMAN, DARBY & CUSHMAN  
 STREET: Eleventh Floor, 1615 L. Street, N.W.  
 CITY: Washington  
 STATE: D.C.  
 COUNTRY: USA  
 ZIP: 20036-5601  
 COMPUTER READABLE FORM:  
 MEDIUM TYPE: Floppy disk  
 COMPUTER: IBM PC compatible  
 OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: PatentIn Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/022,835  
FILING DATE:  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/599,491  
FILING DATE: 17-OCT-1990  
ATTORNEY/AGENT INFORMATION:  
NAME: Scott, Watson T.  
REGISTRATION NUMBER: 26,581  
REFERENCE/DOCKET NUMBER: WTS/5683/78507/SRL  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (202) 861-3000  
TELEFAX: (202) 822-0944  
TELEX: 6714627 CUSH  
INFORMATION FOR SEQ ID NO: 6:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 855 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-08-022-835-6

Query Match 14.6%; Score 57.5; DB 1; Length 855;  
Best Local Similarity 29.5%; Pred. No. 20;  
Matches 23; Conservative 11; Mismatches 27; Indels 17; Gaps 4;

QY 2 IEVVC---NDRLGKVV-----RVKCNDDTIGDLKKLIAAQTGRW----NKIVLKKW-- 47  
DB 296 VEINCTRPNNTRKSIHIGPGRAFTTGTGIIQDINQAHCNLSRAKWNDFLNKIVIKLREQ 355  
QY 48 ---YTIKFDHVSGLDYEI 62  
DB 356 FGNKTIIVFKHSSGGDPEI 373

RESULT 10  
US-08-388-809-6  
Sequence 6, Application US/08388809  
Patent No. 5576000  
GENERAL INFORMATION:  
APPLICANT: REITZ, JR., MARVIN S.; FRANCHINI,  
APPLICANT: GENOVEFFA; MARKHAM, PHILIP D. GALLO, ROBERT  
APPLICANT: C.; LORI, FRANCO C.; POPOVIC, MIKULAS; AND  
APPLICANT: GARTNER, SUZANNE  
TITLE OF INVENTION: MOLECULAR CLONES OF HIV-1  
TITLE OF INVENTION: AND USES THEREOF  
NUMBER OF SEQUENCES: 6  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: MORGAN & FINNEGAN  
STREET: 345 PARK AVENUE  
CITY: NEW YORK  
STATE: NEW YORK  
COUNTRY: USA  
ZIP: 10154  
COMPUTER READABLE FORM:  
MEDIUM TYPE: FLOPPY DISK, 3.5"  
COMPUTER: IBM PC COMPATIBLE  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: WORDPERFECT 5.1  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/388,809  
FILING DATE: 15-FEB-1995  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/022,835  
FILING DATE: 25-FEB-1993  
APPLICATION NUMBER: US 07/599,491  
FILING DATE: 17-OCT-1990  
ATTORNEY/AGENT INFORMATION:  
NAME: LESLIE A. SERUNIAN  
REGISTRATION NUMBER: 35,353

REFERENCE/DOCKET NUMBER: 2026-4092US2  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (212) 758-4800  
TELEFAX: (212) 751-6849  
TELEX: 421792  
INFORMATION FOR SEQ ID NO: 6:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 855 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-08-388-809-6

Query Match 14.6%; Score 57.5; DB 1; Length 855;  
Best Local Similarity 29.5%; Pred. No. 20;  
Matches 23; Conservative 11; Mismatches 27; Indels 17; Gaps 4;

QY 2 IEVVC---NDRLGKVV-----RVKCNDDTIGDLKKLIAAQTGRW----NKIVLKKW-- 47  
DB 296 VEINCTRPNNTRKSIHIGPGRAFTTGTGIIQDINQAHCNLSRAKWNDFLNKIVIKLREQ 355  
QY 48 ---YTIKFDHVSGLDYEI 62  
DB 356 FGNKTIIVFKHSSGGDPEI 373

RESULT 11  
US-08-647-714-6  
Sequence 6, Application US/08647714  
Patent No. 5869313  
GENERAL INFORMATION:  
APPLICANT: REITZ, JR., MARVIN S.; FRANCHINI,  
APPLICANT: GENOVEFFA; MARKHAM, PHILIP D. GALLO, ROBERT  
APPLICANT: C.; LORI, FRANCO C.; POPOVIC, MIKULAS; AND  
APPLICANT: GARTNER, SUZANNE  
TITLE OF INVENTION: MOLECULAR CLONES OF HIV-1  
TITLE OF INVENTION: AND USES THEREOF  
NUMBER OF SEQUENCES: 6  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: MORGAN & FINNEGAN  
STREET: 345 PARK AVENUE  
CITY: NEW YORK  
STATE: NEW YORK  
COUNTRY: USA  
ZIP: 10154  
COMPUTER READABLE FORM:  
MEDIUM TYPE: FLOPPY DISK, 3.5"  
COMPUTER: IBM PC COMPATIBLE  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: WORDPERFECT 5.1  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/647,714  
FILING DATE:  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US/08/388,809  
FILING DATE: 15-FEB-1995  
APPLICATION NUMBER: US 08/022,835  
FILING DATE: 25-FEB-1993  
APPLICATION NUMBER: US 07/599,491  
FILING DATE: 17-OCT-1990  
ATTORNEY/AGENT INFORMATION:  
NAME: LESLIE A. SERUNIAN  
REGISTRATION NUMBER: 35,353  
REFERENCE/DOCKET NUMBER: 2026-4092US2  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (212) 758-4800  
TELEFAX: (212) 751-6849  
TELEX: 421792  
INFORMATION FOR SEQ ID NO: 6:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 855 amino acids  
TYPE: amino acid

: PATENT NO.: 3603624  
 : GENERAL INFORMATION:  
 : APPLICANT: Hasel, Karl W.  
 : APPLICANT: Maddon, Paul J.  
 : TITLE OF INVENTION: HIV-1  
 : TITLE OF INVENTION: THERE  
 : TITLE OF INVENTION: THERE  
 : NUMBER OF SEQUENCES: 29  
 :

CORRESPONDENCE ADDRESS:  
ADDRESSEE: Cooper & Dunham  
STREET: 30 Rockefeller Plaza  
CITY: New York  
STATE: New York  
COUNTRY: USA  
ZIP: 10112  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.24  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/037,816A  
FILING DATE: 26-MAR-1993  
CLASSIFICATION: 435  
ATTORNEY/AGENT INFORMATION:  
NAME: White, John P.  
REGISTRATION NUMBER: 28 678  
REFERENCE/DOCKET NUMBER: 41190/JPW/AJMJ  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (212) 977-9550  
TELEFAX: (212) 664-0525  
TELEX: 422523 COOPUI  
INFORMATION FOR SEQ ID NO: 28:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 507 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-08-037-816A-28

[illegible]

RESULT 15  
US-08-530146-16  
; Sequence 16, Application US/08530146  
; Patent No. 5886163  
; GENERAL INFORMATION:  
; APPLICANT: Hasel, Karl W.  
; APPLICANT: Maddon, Paul J.  
; TITLE OF INVENTION: HIV-1 VACCINES, ANTIBODY COMPOSITIONS RELATED  
; TITLE OF INVENTION: THERETO, AND THERAPEUTIC AND PROPHYLACTIC USES  
; TITLE OF INVENTION: THEREOF  
; NUMBER OF SEQUENCES: 29  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Cooper & Dunham  
; STREET: 30 Rockefeller Plaza  
; CITY: New York  
; STATE: New York  
; COUNTRY: USA  
; ZIP: 10112  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC Compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patent in Release #1.24  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/530,146  
; FILING DATE:  
; CLASSIFICATION:

PRIORITY APPLICATION DATA:  
 APPLICATION NUMBER: US/08/037,816  
 FILING DATE: 26-MAR-1993  
 ATTORNEY/AGENT INFORMATION:  
 NAME: White, John P.  
 REGISTRATION NUMBER: 28,678  
 REFERENCE/DOCKET NUMBER: 41190/JPW/AJMJ  
 TELECOMMUNICATION INFORMATION:  
 TELEPHONE: (212) 977-9550  
 TELEFAX: (212) 664-0525  
 TELEX: 422523 COOPUI  
 INFORMATION FOR SEQ ID NO: 16:  
 SEQUENCE CHARACTERISTICS:  
 LENGTH: 507 amino acids  
 TYPE: amino acid  
 TOPOLOGY: linear  
 MOLECULE TYPE: protein  
 US-08-530-146-16

```

Query Match      14.3%; Score 56.5; DB 2; Length 507;
Best Local Similarity 28.2%; Pred. No. 14;
Matches 22; Conservative 14; Mismatches 25; Indels 17; Gaps 4;

QY      2 IEVVC---NDRLGKKV-----RVKCNTDGDTIGDLKKLIAAGTGFWN----KIVLK-----45
       ||| | | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
Db      295 VEINCTRPNNTRKSIHIGPGRAFYTTGEIIGDIRQAHCNISRAKWNDLTQIKVIKLREQ 354

QY      46 -KWTTFKDHSVLSGDYEI 62
       : || : || || || ||
Db      355 FENKTIVFNSSGGDPEI 372
```

Search completed: June 13, 2001, 20:26:08  
Job time: 1626 sec

GenCore version 4.5  
Copyright (C) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: June 13, 2001, 19:31:27 ; Search time 59.54 Seconds  
(without alignments)  
70.086 Million cell updates/sec

Title: US-09-331-930A-2  
Perfect score: 394  
Sequence: 1 MIEVVNDRLGKVKVRCNT.....HVSGLDYEIHDMNLELYIQ 73

Scoring table:  
BLOSUM62  
Gapop 10.0 , Gapext 0.5

Archived: 390729 seqs, 57163235 residues

Total number of hits satisfying chosen parameters: 390729

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000  
Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : A\_Geneseq\_0401.\*  
1: /SID56/gcgdata/geneseq/geneseq/AA1980.DAT.\*  
2: /SID56/gcgdata/geneseq/geneseq/AA1981.DAT.\*  
3: /SID56/gcgdata/geneseq/geneseq/AA1982.DAT.\*  
4: /SID56/gcgdata/geneseq/geneseq/AA1983.DAT.\*  
5: /SID56/gcgdata/geneseq/geneseq/AA1984.DAT.\*  
6: /SID56/gcgdata/geneseq/geneseq/AA1985.DAT.\*  
7: /SID56/gcgdata/geneseq/geneseq/AA1986.DAT.\*  
8: /SID56/gcgdata/geneseq/geneseq/AA1987.DAT.\*  
9: /SID56/gcgdata/geneseq/geneseq/AA1988.DAT.\*  
10: /SID56/gcgdata/geneseq/geneseq/AA1989.DAT.\*  
11: /SID56/gcgdata/geneseq/geneseq/AA1990.DAT.\*  
12: /SID56/gcgdata/geneseq/geneseq/AA1991.DAT.\*  
13: /SID56/gcgdata/geneseq/geneseq/AA1992.DAT.\*  
14: /SID56/gcgdata/geneseq/geneseq/AA1993.DAT.\*  
15: /SID56/gcgdata/geneseq/geneseq/AA1994.DAT.\*  
16: /SID56/gcgdata/geneseq/geneseq/AA1995.DAT.\*  
17: /SID56/gcgdata/geneseq/geneseq/AA1996.DAT.\*  
18: /SID56/gcgdata/geneseq/geneseq/AA1997.DAT.\*  
19: /SID56/gcgdata/geneseq/geneseq/AA1998.DAT.\*  
20: /SID56/gcgdata/geneseq/geneseq/AA1999.DAT.\*  
21: /SID56/gcgdata/geneseq/geneseq/AA2000.DAT.\*  
22: /SID56/gcgdata/geneseq/geneseq/AA2001.DAT.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match %	Length	ID	Description
1	394	100.0	73	20 Y08413	P. obesus beaon p
2	394	100.0	73	21 B36290	Israeli sand rat b
3	313	79.4	73	21 G41925	Arabidopsis thalia
4	313	79.4	96	21 G41924	Arabidopsis thalia
5	312	79.2	73	21 G40885	zea mays protein f
6	300	76.1	73	21 G27281	zea mays protein f
7	163	41.4	33	20 Y08414	Human beaon prote
8	147.5	37.4	32	21 B36291	Human beaon (shor
9	80	20.3	477	21 G38562	Arabidopsis thalia
10	80	20.3	553	21 G38561	Arabidopsis thalia
11	80	20.3	631	21 G38560	Arabidopsis thalia

12	67	17.0	236	21	B41138	Human OREF902
13	60	15.2	289	21	B42562	Human OREF902
14	60	15.2	295	21	G23780	Arabidopsis thalia
15	60	15.2	296	21	G23779	Arabidopsis thalia
16	60	15.2	323	21	G23778	Arabidopsis thalia
17	60	15.2	494	19	W37060	HIV-1 breakthrough
18	59.5	15.1	513	20	W97587	Human immunodefici
19	59	15.0	455	19	W37069	HIV-1 gp120 MN-GNE
20	59	15.0	687	22	B49465	Canine retrovirus
21	59	15.0	850	16	R67724	gp120 from the HIV
22	58.5	14.8	365	21	B18259	Plasmodium falcipa
23	58.5	14.8	498	19	W37054	HIV-1 breakthrough
24	58.5	14.8	498	19	W37055	HIV-1 breakthrough
25	58.5	14.8	855	21	Y96946	HIV synthetic Env
26	58	14.7	483	19	W37064	HIV-1 breakthrough
27	58	14.7	487	19	W37065	HIV-1 breakthrough
28	58	14.7	822	19	W69847	Amino acid sequenc
29	57.5	14.6	368	21	G44390	Arabidopsis thalia
30	57.5	14.6	465	21	G44389	Arabidopsis thalia
31	57.5	14.6	498	21	G44388	Arabidopsis thalia
32	57.5	14.6	660	21	B54249	Human pancreatic c
33	57.5	14.6	662	20	Y05377	Human HCMV Inducib
34	57.5	14.6	855	12	R14905	HIV-1 BA-L clone-e
35	57.5	14.6	855	18	W11581	Human Immunodefici
36	57.5	14.6	855	20	W88113	Env protein of the
37	57	14.5	97	21	G15024	Arabidopsis thalia
38	57	14.5	139	21	G15023	Arabidopsis thalia
39	56.5	14.3	132	21	G44405	Arabidopsis thalia
40	56.5	14.3	222	17	R90670	S. cerevisiae delt
41	56.5	14.3	507	15	R60783	HIV virus-1(JR-FL)
42	56.5	14.3	507	15	R60789	HIV virus-1(JR-FL)
43	56.5	14.3	849	21	Y96945	HIV synthetic Env
44	56	14.2	88	21	G15025	Arabidopsis thalia
45	56	14.2	270	19	W41250	Xenopus cerberus p

## ALIGNMENTS

RESULT 1	
Y08413	
ID Y08413 standard; Protein; 73 AA.	
XX	
AC Y08413;	
XX	
DT 24-JUL-1999 (first entry)	
XX	
DE P. obesus beaon protein.	
XX	
KW Beacon; hypothalamus; obese; lean; agonist; antagonist; treatment;	
KW obesity; anorexia; weight maintenance; energy imbalance; diabetes;	
KW metabolic syndrome; dyslipidemia; hypertension; insulin resistance;	
KW medicament; livestock; diagnosis.	
XX	
OS Psammomys obesus.	
XX	
PN WO9923217-A1.	
XX	
PD 14-MAY-1999.	
XX	
PF 30-OCT-1998; 98WO-AU00902.	
XX	
PR 11-NOV-1997; 97AU-0000323.	
PR 31-OCT-1997; 97AU-0000117.	
XX	
PA (ITDI-) INT DIABETES INST.	
PA (UYDE-) UNIV DEAKIN.	
XX	
PI Collier G, Zimmert PZ;	
XX	
DR WPI; 1999-337484/28.	
DR N-PSDB; X57359.	
XX	

PT New gene encoding a beacon protein associated with modulation of  
XX obesity, diabetes and metabolic energy levels  
XX Claim 2; Page 50; 85pp; English.

XX This invention describes a novel beacon protein and its encoding nucleic  
CC acid which is expressed in larger amounts in hypothalamus tissue of obese  
CC animals compared to lean animals. Agonists and antagonists of beacon can  
CC be used to treat obesity, anorexia, weight maintenance, energy imbalance,  
CC diabetes, metabolic syndrome, dyslipidemia, hypertension and/or insulin  
CC resistance. The beacon protein, itself is used to manufacture medicaments  
CC for treatment of obesity, anorexia, energy imbalance or diabetes. The  
CC treatment is contemplated for both human and animals, such as those  
CC important to the livestock industry. The antibody and polynucleotides are  
CC useful in diagnosis of conditions as above.

XX Sequence 73 AA;

Query Match 100.0%; Score 394; DB 20; Length 73;  
Best Local Similarity 100.0%; Pred. No. 1.2e-42;  
Matches 73; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 1 MIEVVCNDRLGKKVRVKNCTDDTIGDLKKLIAAQTGRNKKIVLKKWYTFKDHVSLG DY 60  
|||||  
1 mievvcndrlgkvrkvkcntddtldgdlkklliaaqtgrwnkivlkkwtyfkdhvslgdy 60  
Oy 61 EIHGGMNLELYQ 73  
|||||  
Db 61 eihdgmmllelyyq 73

RESULT 2

ID B36290

AC B36290 standard; Protein; 73 AA.

XX B36290;

XX 23-FEB-2001 (first entry)

XX Israeli sand rat beacon ligand.

XX Israeli sand rat; beacon; obesity; NIDDM; energy balance; diabetes;

XX Psammomys obesus.

XX WO200064931-A1.

XX 02-NOV-2000.

XX 19-APR-2000; 2000WO-AU00342.

XX 23-APR-1999; 99AU-00099919.

XX 24-MAR-2000; 2000AU-0006454.

XX (AUTO-) AUTOGEN PTY LTD.

XX Collier G, Walder K, Zimmet P;

XX WPI; 2000-687311/67.

XX N-PSDB; C81767.

XX Ligand of beacon protein useful for treating obesity, anorexia, energy  
XX imbalance, diabetes, metabolic syndrome, dyslipidemia, hypertension and  
XX insulin resistance -

XX Claim 3; Fig 1; 67pp; English.

XX The present invention is related to the isolation of a ligand known as  
XX beacon from the Israeli sand rat. Beacon is associated with the  
XX regulation of energy balance, and the protein, its coding sequence and  
XX analogues can be used in the treatment of diabetes, obesity, anorexia,

CC energy imbalance, metabolic syndrome, dyslipidaemia, hypertension and  
CC insulin resistance. In addition, they can be used in agriculture to  
XX produce leaner animals.  
XX Sequence 73 AA;

Query Match 100.0%; Score 394; DB 21; Length 73;  
Best Local Similarity 100.0%; Pred. No. 1.2e-42;  
Matches 73; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 1 MIEVVCNDRLGKKVRVKNCTDDTIGDLKKLIAAQTGRNKKIVLKKWYTFKDHVSLG DY 60  
|||||  
Db 1 mievvcndrlgkvrkvkcntddtldgdlkklliaaqtgrwnkivlkkwtyfkdhvslgdy 60  
Oy 61 EIHGGMNLELYQ 73  
|||||  
Db 61 eihdgmmllelyyq 73

RESULT 3

G41925

ID G41925 standard; Protein; 73 AA.

XX G41925;

XX 18-OCT-2000 (first entry)

XX Arabidopsis thaliana protein fragment SEQ ID NO: 52222.

XX Protein identification; signal transduction pathway; metabolic pathway;

XX hybridisation assay; genetic mapping; gene expression control; promoter;

XX termination sequence.

XX Arabidopsis thaliana.

XX EP1033405-A2.

XX 06-SEP-2000.

XX 25-FEB-2000; 2000EP-0301439.

XX 25-FEB-1999; 99US-0121825.

XX 05-MAR-1999; 99US-0123180.

XX 09-MAR-1999; 99US-0123548.

XX 23-MAR-1999; 99US-0125788.

XX 25-MAR-1999; 99US-0126264.

XX 29-MAR-1999; 99US-0126785.

XX 01-APR-1999; 99US-0127462.

XX 06-APR-1999; 99US-0128234.

XX 16-APR-1999; 99US-0129845.

XX 19-APR-1999; 99US-0130077.

XX 21-APR-1999; 99US-0130449.

XX 23-APR-1999; 99US-0130510.

XX 28-APR-1999; 99US-0130891.

XX 30-APR-1999; 99US-0131449.

XX 30-APR-1999; 99US-0132048.

XX 30-APR-1999; 99US-0132407.

XX 04-MAY-1999; 99US-0132484.

XX 05-MAY-1999; 99US-0132485.

XX 06-MAY-1999; 99US-0132486.

XX 06-MAY-1999; 99US-0132487.

XX 07-MAY-1999; 99US-0132863.

XX 11-MAY-1999; 99US-0134256.

XX 14-MAY-1999; 99US-0134218.

XX 14-MAY-1999; 99US-0134219.

XX 14-MAY-1999; 99US-0134221.

XX 14-MAY-1999; 99US-0134370.

XX 18-MAY-1999; 99US-0134768.

XX 19-MAY-1999; 99US-0134941.

XX 20-MAY-1999; 99US-0135124.

XX 21-MAY-1999; 99US-0135353.

PR 24-MAY-1999; 99US-0135629.  
PR 25-MAY-1999; 99US-0136021.  
PR 27-MAY-1999; 99US-0136392.  
PR 28-MAY-1999; 99US-0136782.  
PR 01-JUN-1999; 99US-0137222.  
PR 03-JUN-1999; 99US-0137528.  
PR 04-JUN-1999; 99US-0137502.  
PR 07-JUN-1999; 99US-0137724.  
PR 08-JUN-1999; 99US-0138094.  
PR 10-JUN-1999; 99US-0138540.  
PR 10-JUN-1999; 99US-0138847.  
PR 14-JUN-1999; 99US-0139119.  
PR 16-JUN-1999; 99US-0139452.  
PR 17-JUN-1999; 99US-0139453.  
PR 17-JUN-1999; 99US-0139492.  
PR 18-JUN-1999; 99US-0139454.  
PR 18-JUN-1999; 99US-0139455.  
PR 18-JUN-1999; 99US-0139456.  
PR 18-JUN-1999; 99US-0139457.  
PR 18-JUN-1999; 99US-0139458.  
PR 18-JUN-1999; 99US-0139459.  
PR 18-JUN-1999; 99US-0139460.  
PR 18-JUN-1999; 99US-0139461.  
PR 18-JUN-1999; 99US-0139462.  
PR 18-JUN-1999; 99US-0139463.  
PR 18-JUN-1999; 99US-0139750.  
PR 18-JUN-1999; 99US-0139763.  
PR 21-JUN-1999; 99US-0139817.  
PR 22-JUN-1999; 99US-0139899.  
PR 23-JUN-1999; 99US-0140353.  
PR 23-JUN-1999; 99US-0140354.  
PR 24-JUN-1999; 99US-0140695.  
PR 28-JUN-1999; 99US-0140823.  
PR 29-JUN-1999; 99US-0140991.  
PR 30-JUN-1999; 99US-0141287.  
PR 01-JUL-1999; 99US-0141842.  
PR 01-JUL-1999; 99US-0142154.  
PR 02-JUL-1999; 99US-0142055.  
PR 06-JUL-1999; 99US-0142390.  
PR 08-JUL-1999; 99US-0142803.  
PR 09-JUL-1999; 99US-0142920.  
PR 12-JUL-1999; 99US-0142977.  
PR 13-JUL-1999; 99US-0143542.  
PR 14-JUL-1999; 99US-0143624.  
PR 15-JUL-1999; 99US-0144005.  
PR 16-JUL-1999; 99US-0144085.  
PR 16-JUL-1999; 99US-0144086.  
PR 19-JUL-1999; 99US-0144325.  
PR 19-JUL-1999; 99US-0144331.  
PR 19-JUL-1999; 99US-0144332.  
PR 19-JUL-1999; 99US-0144333.  
PR 19-JUL-1999; 99US-0144334.  
PR 19-JUL-1999; 99US-0144335.  
PR 20-JUL-1999; 99US-0144352.  
PR 20-JUL-1999; 99US-0144632.  
PR 20-JUL-1999; 99US-0144884.  
PR 21-JUL-1999; 99US-0144814.  
PR 21-JUL-1999; 99US-0145086.  
PR 21-JUL-1999; 99US-0145088.  
PR 22-JUL-1999; 99US-0145085.  
PR 22-JUL-1999; 99US-0145087.  
PR 22-JUL-1999; 99US-0145089.  
PR 22-JUL-1999; 99US-0145192.  
PR 23-JUL-1999; 99US-0145145.  
PR 23-JUL-1999; 99US-0145218.  
PR 23-JUL-1999; 99US-0145224.  
PR 26-JUL-1999; 99US-0145276.  
PR 27-JUL-1999; 99US-0145913.  
PR 27-JUL-1999; 99US-0145918.  
PR 28-JUL-1999; 99US-0145919.  
PR 28-JUL-1999; 99US-0145951.  
PR 02-AUG-1999; 99US-0146386.  
PR 02-AUG-1999; 99US-0146388.  
PR 02-AUG-1999; 99US-0146389.  
PR 03-AUG-1999; 99US-0147038.  
PR 04-AUG-1999; 99US-0147204.  
PR 04-AUG-1999; 99US-0147302.  
PR 05-AUG-1999; 99US-0147192.  
PR 05-AUG-1999; 99US-0147260.  
PR 06-AUG-1999; 99US-0147303.  
PR 06-AUG-1999; 99US-0147416.  
PR 09-AUG-1999; 99US-0147493.  
PR 09-AUG-1999; 99US-0147935.  
PR 10-AUG-1999; 99US-0148171.  
PR 11-AUG-1999; 99US-0148319.  
PR 12-AUG-1999; 99US-0148341.  
PR 13-AUG-1999; 99US-0148565.  
PR 13-AUG-1999; 99US-0148684.  
PR 16-AUG-1999; 99US-0149368.  
PR 17-AUG-1999; 99US-0149175.  
PR 18-AUG-1999; 99US-0149426.  
PR 20-AUG-1999; 99US-0149722.  
PR 20-AUG-1999; 99US-0149723.  
PR 20-AUG-1999; 99US-0149929.  
PR 23-AUG-1999; 99US-0149902.  
PR 23-AUG-1999; 99US-0149930.  
PR 25-AUG-1999; 99US-0150566.  
PR 26-AUG-1999; 99US-0150884.  
PR 27-AUG-1999; 99US-0151065.  
PR 27-AUG-1999; 99US-0151066.  
PR 27-AUG-1999; 99US-0151080.  
PR 30-AUG-1999; 99US-0151303.  
PR 31-AUG-1999; 99US-0151438.  
PR 01-SEP-1999; 99US-0151930.  
PR 07-SEP-1999; 99US-0152363.  
PR 10-SEP-1999; 99US-0153070.  
PR 13-SEP-1999; 99US-0153758.  
PR 13-SEP-1999; 99US-0154018.  
PR 16-SEP-1999; 99US-0154039.  
PR 20-SEP-1999; 99US-0154779.  
PR 22-SEP-1999; 99US-0155139.  
PR 23-SEP-1999; 99US-0155486.  
PR 24-SEP-1999; 99US-0155659.  
PR 28-SEP-1999; 99US-0156458.  
PR 29-SEP-1999; 99US-0156596.  
PR 04-OCT-1999; 99US-0157117.  
PR 05-OCT-1999; 99US-0157753.  
PR 06-OCT-1999; 99US-0157865.  
PR 07-OCT-1999; 99US-0158029.  
PR 08-OCT-1999; 99US-0158232.  
PR 12-OCT-1999; 99US-0158369.  
PR 13-OCT-1999; 99US-0159293.  
PR 13-OCT-1999; 99US-0159294.  
PR 13-OCT-1999; 99US-0159295.  
PR 14-OCT-1999; 99US-0159329.  
PR 14-OCT-1999; 99US-0159330.  
PR 14-OCT-1999; 99US-0159331.  
PR 14-OCT-1999; 99US-0159637.  
PR 14-OCT-1999; 99US-0159638.  
PR 18-OCT-1999; 99US-0159584.  
PR 21-OCT-1999; 99US-0160741.  
PR 21-OCT-1999; 99US-0160767.  
PR 21-OCT-1999; 99US-0160768.  
PR 21-OCT-1999; 99US-0160770.  
PR 21-OCT-1999; 99US-0160814.  
PR 21-OCT-1999; 99US-0160815.  
PR 22-OCT-1999; 99US-0160980.  
PR 22-OCT-1999; 99US-0160981.  
PR 22-OCT-1999; 99US-0160989.  
PR 25-OCT-1999; 99US-0161404.  
PR 25-OCT-1999; 99US-0161405.  
PR 25-OCT-1999; 99US-0161406.  
PR 26-OCT-1999; 99US-0161359.  
PR 26-OCT-1999; 99US-0161360.  
PR 26-OCT-1999; 99US-0161361.  
PR 28-OCT-1999; 99US-0161920.





```
PR 04-AUG-1999; 99US-0147302.
PR 05-AUG-1999; 99US-0147192.
PR 05-AUG-1999; 99US-0147260.
PR 06-AUG-1999; 99US-0147303.
PR 06-AUG-1999; 99US-0147416.
PR 09-AUG-1999; 99US-0147493.
PR 09-AUG-1999; 99US-0147935.
PR 10-AUG-1999; 99US-0148171.
PR 11-AUG-1999; 99US-0148319.
PR 12-AUG-1999; 99US-0148341.
PR 13-AUG-1999; 99US-0148565.
PR 13-AUG-1999; 99US-0148684.
PR 16-AUG-1999; 99US-0149368.
PR 17-AUG-1999; 99US-0149175.
PR 18-AUG-1999; 99US-0149426.
PR 20-AUG-1999; 99US-0149722.
PR 20-AUG-1999; 99US-0149723.
PR 20-AUG-1999; 99US-0149929.
PR 23-AUG-1999; 99US-0149902.
PR 23-AUG-1999; 99US-0149930.
PR 25-AUG-1999; 99US-0150566.
PR 26-AUG-1999; 99US-0150884.
PR 27-AUG-1999; 99US-0151065.
PR 27-AUG-1999; 99US-0151066.
PR 30-AUG-1999; 99US-0151080.
PR 30-AUG-1999; 99US-0151303.
PR 31-AUG-1999; 99US-0151438.
PR 01-SEP-1999; 99US-0151930.
PR 07-SEP-1999; 99US-0152363.
PR 10-SEP-1999; 99US-0153070.
PR 13-SEP-1999; 99US-0153758.
PR 15-SEP-1999; 99US-0154018.
PR 16-SEP-1999; 99US-0154039.
PR 20-SEP-1999; 99US-0154779.
PR 22-SEP-1999; 99US-0155139.
PR 23-SEP-1999; 99US-0155486.
PR 24-SEP-1999; 99US-0155659.
PR 28-SEP-1999; 99US-0156458.
PR 29-SEP-1999; 99US-0156596.
PR 04-OCT-1999; 99US-0157117.
PR 05-OCT-1999; 99US-0157753.
PR 06-OCT-1999; 99US-0157865.
PR 07-OCT-1999; 99US-0158029.
PR 08-OCT-1999; 99US-0158232.
PR 12-OCT-1999; 99US-0158369.
PR 13-OCT-1999; 99US-0159293.
PR 13-OCT-1999; 99US-0159294.
PR 13-OCT-1999; 99US-0159298.
PR 14-OCT-1999; 99US-0159329.
PR 14-OCT-1999; 99US-0159330.
PR 14-OCT-1999; 99US-0159331.
PR 14-OCT-1999; 99US-0159637.
PR 14-OCT-1999; 99US-0159638.
PR 18-OCT-1999; 99US-0159584.
PR 21-OCT-1999; 99US-0160741.
PR 21-OCT-1999; 99US-0160767.
PR 21-OCT-1999; 99US-0160768.
PR 21-OCT-1999; 99US-0160770.
PR 21-OCT-1999; 99US-0160814.
PR 21-OCT-1999; 99US-0160815.
PR 22-OCT-1999; 99US-0160980.
PR 22-OCT-1999; 99US-0160981.
PR 22-OCT-1999; 99US-0160989.
PR 25-OCT-1999; 99US-0161404.
PR 25-OCT-1999; 99US-0161405.
PR 25-OCT-1999; 99US-0161406.
PR 26-OCT-1999; 99US-0161359.
PR 26-OCT-1999; 99US-0161360.
PR 26-OCT-1999; 99US-0161361.
PR 28-OCT-1999; 99US-0161920.
PR 28-OCT-1999; 99US-0161992.
PR 28-OCT-1999; 99US-0161993.
PR 29-OCT-1999; 99US-0162142.

Query Match 79.4%; Score 313; DB 21; Length 96;
Best Local Similarity 80.6%; Pred. No. 3e-32;
Matches 58; Conservative 6; Mismatches 8; Indels 0; Gaps 0;

QY 1 MIEVVCNDRLGKKVRKVCNTDDTTGDLKLLIAAOTGTWNNKIVLKKWYTIKFDHVS LGDY 60
Db 24 mievvlndrigrkvrkvcndddttgdlkllvaaqtgrae kirkqkwniykdhitlkdy 83
QY 61 EIHGDMNLELYY 72
Db 84 eihdgmglelyy 95

RESULT 5
G40885
ID G40885 standard; Protein; 73 AA.
XX
AC G40885;
XX
DT 18-OCT-2000 (first entry)
XX
DE Zea mays protein fragment SEQ ID NO: 50790.
XX
KW Protein identification; signal transduction pathway; metabolic pathway;
KW hybridisation assay; genetic mapping; gene expression control; promoter;
XX termination sequence; corn.
XX
OS Zea mays subsp. mays.
XX
PN EP1033405-A2.
XX
PD 06-SEP-2000.
XX
PF 25-FEB-2000; 2000EP-0301439.
XX
PR 25-FEB-1999; 99US-0121825.
PR 05-MAR-1999; 99US-0123180.
PR 09-MAR-1999; 99US-0123548.
PR 23-MAR-1999; 99US-0125788.
PR 25-MAR-1999; 99US-0126264.
PR 29-MAR-1999; 99US-0126785.
PR 01-APR-1999; 99US-0127462.
PR 06-APR-1999; 99US-0128234.
PR 08-APR-1999; 99US-0128714.
PR 16-APR-1999; 99US-0129845.
PR 19-APR-1999; 99US-0130077.
PR 21-APR-1999; 99US-0130449.
PR 23-APR-1999; 99US-0130510.
PR 23-APR-1999; 99US-0130891.
PR 28-APR-1999; 99US-0131449.
PR 30-APR-1999; 99US-0132048.
PR 30-APR-1999; 99US-0132407.
PR 04-MAY-1999; 99US-0132484.
PR 05-MAY-1999; 99US-0132485.
PR 06-MAY-1999; 99US-0132486.
PR 06-MAY-1999; 99US-0132487.
PR 07-MAY-1999; 99US-0132863.
PR 11-MAY-1999; 99US-0134256.
PR 14-MAY-1999; 99US-0134218.
PR 14-MAY-1999; 99US-0134219.
PR 14-MAY-1999; 99US-0134221.
PR 14-MAY-1999; 99US-0134370.
PR 18-MAY-1999; 99US-0134768.
PR 19-MAY-1999; 99US-0134941.
PR 20-MAY-1999; 99US-0135124.
PR 21-MAY-1999; 99US-0135353.
PR 24-MAY-1999; 99US-0135629.
PR 25-MAY-1999; 99US-0136021.
PR 27-MAY-1999; 99US-0136392.
PR 28-MAY-1999; 99US-0136782.
PR 01-JUN-1999; 99US-0137222.
PR 03-JUN-1999; 99US-0137528.
```

PR	04-JUN-1999;	99US-0137502.	PR	06-AUG-1999;	99US-0147303.
PR	07-JUN-1999;	99US-0137724.	PR	06-AUG-1999;	99US-0147416.
PR	08-JUN-1999;	99US-0138094.	PR	09-AUG-1999;	99US-0147493.
PR	10-JUN-1999;	99US-0138540.	PR	09-AUG-1999;	99US-0147935.
PR	10-JUN-1999;	99US-0138847.	PR	10-AUG-1999;	99US-0148171.
PR	14-JUN-1999;	99US-0139119.	PR	11-AUG-1999;	99US-0148319.
PR	16-JUN-1999;	99US-0139452.	PR	12-AUG-1999;	99US-0148341.
PR	16-JUN-1999;	99US-0139453.	PR	13-AUG-1999;	99US-0148565.
PR	17-JUN-1999;	99US-0139492.	PR	13-AUG-1999;	99US-0148684.
PR	18-JUN-1999;	99US-0139454.	PR	16-AUG-1999;	99US-0149368.
PR	18-JUN-1999;	99US-0139455.	PR	17-AUG-1999;	99US-0149175.
PR	18-JUN-1999;	99US-0139456.	PR	18-AUG-1999;	99US-0149426.
PR	18-JUN-1999;	99US-0139457.	PR	20-AUG-1999;	99US-0149722.
PR	18-JUN-1999;	99US-0139458.	PR	20-AUG-1999;	99US-0149723.
PR	18-JUN-1999;	99US-0139459.	PR	20-AUG-1999;	99US-0149929.
PR	18-JUN-1999;	99US-0139460.	PR	23-AUG-1999;	99US-0149902.
PR	18-JUN-1999;	99US-0139461.	PR	23-AUG-1999;	99US-0149930.
PR	18-JUN-1999;	99US-0139463.	PR	25-AUG-1999;	99US-0150566.
PR	18-JUN-1999;	99US-0139463.	PR	26-AUG-1999;	99US-0150884.
PR	18-JUN-1999;	99US-0139750.	PR	27-AUG-1999;	99US-0151065.
PR	18-JUN-1999;	99US-0139763.	PR	27-AUG-1999;	99US-0151066.
PR	21-JUN-1999;	99US-0139817.	PR	27-AUG-1999;	99US-0151080.
PR	22-JUN-1999;	99US-0139899.	PR	30-AUG-1999;	99US-0151303.
PR	23-JUN-1999;	99US-0140353.	PR	31-AUG-1999;	99US-0151438.
PR	23-JUN-1999;	99US-0140354.	PR	01-SEP-1999;	99US-0151930.
PR	24-JUN-1999;	99US-0140895.	PR	07-SEP-1999;	99US-0152363.
PR	28-JUN-1999;	99US-0140823.	PR	10-SEP-1999;	99US-0153070.
PR	29-JUN-1999;	99US-0140991.	PR	13-SEP-1999;	99US-0153758.
PR	30-JUN-1999;	99US-0141287.	PR	15-SEP-1999;	99US-0154018.
PR	01-JUL-1999;	99US-0141842.	PR	16-SEP-1999;	99US-0154039.
PR	01-JUL-1999;	99US-0142154.	PR	20-SEP-1999;	99US-0154779.
PR	02-JUL-1999;	99US-0142055.	PR	22-SEP-1999;	99US-0155139.
PR	06-JUL-1999;	99US-0142390.	PR	23-SEP-1999;	99US-0155486.
PR	08-JUL-1999;	99US-0142803.	PR	24-SEP-1999;	99US-0155659.
PR	09-JUL-1999;	99US-0142920.	PR	28-SEP-1999;	99US-0156458.
PR	12-JUL-1999;	99US-0142977.	PR	29-SEP-1999;	99US-0156596.
PR	13-JUL-1999;	99US-0143542.	PR	04-OCT-1999;	99US-0157117.
PR	14-JUL-1999;	99US-0143624.	PR	05-OCT-1999;	99US-0157753.
PR	15-JUL-1999;	99US-0144005.	PR	06-OCT-1999;	99US-0157865.
PR	16-JUL-1999;	99US-0144085.	PR	07-OCT-1999;	99US-0158029.
PR	16-JUL-1999;	99US-0144086.	PR	08-OCT-1999;	99US-0158232.
PR	19-JUL-1999;	99US-0144325.	PR	12-OCT-1999;	99US-0158369.
PR	19-JUL-1999;	99US-0144331.	PR	13-OCT-1999;	99US-0159293.
PR	19-JUL-1999;	99US-0144332.	PR	13-OCT-1999;	99US-0159294.
PR	19-JUL-1999;	99US-0144333.	PR	13-OCT-1999;	99US-0159295.
PR	19-JUL-1999;	99US-0144334.	PR	14-OCT-1999;	99US-0159329.
PR	19-JUL-1999;	99US-0144335.	PR	14-OCT-1999;	99US-0159330.
PR	20-JUL-1999;	99US-0144352.	PR	14-OCT-1999;	99US-0159331.
PR	20-JUL-1999;	99US-0144632.	PR	14-OCT-1999;	99US-0159637.
PR	20-JUL-1999;	99US-0144884.	PR	14-OCT-1999;	99US-0159638.
PR	21-JUL-1999;	99US-0144814.	PR	18-OCT-1999;	99US-0159584.
PR	21-JUL-1999;	99US-0145086.	PR	21-OCT-1999;	99US-0160741.
PR	21-JUL-1999;	99US-0145088.	PR	21-OCT-1999;	99US-0160767.
PR	22-JUL-1999;	99US-0145085.	PR	21-OCT-1999;	99US-0160768.
PR	22-JUL-1999;	99US-0145087.	PR	21-OCT-1999;	99US-0160770.
PR	22-JUL-1999;	99US-0145089.	PR	21-OCT-1999;	99US-0160814.
PR	22-JUL-1999;	99US-0145192.	PR	21-OCT-1999;	99US-0160815.
PR	23-JUL-1999;	99US-0145145.	PR	22-OCT-1999;	99US-0160980.
PR	23-JUL-1999;	99US-0145218.	PR	22-OCT-1999;	99US-0160981.
PR	23-JUL-1999;	99US-0145224.	PR	22-OCT-1999;	99US-0160989.
PR	26-JUL-1999;	99US-0145276.	PR	25-OCT-1999;	99US-0161404.
PR	27-JUL-1999;	99US-0145913.	PR	25-OCT-1999;	99US-0161405.
PR	27-JUL-1999;	99US-0145918.	PR	25-OCT-1999;	99US-0161406.
PR	28-JUL-1999;	99US-0145919.	PR	26-OCT-1999;	99US-0161359.
PR	28-JUL-1999;	99US-0145951.	PR	26-OCT-1999;	99US-0161360.
PR	02-AUG-1999;	99US-0146386.	PR	26-OCT-1999;	99US-0161361.
PR	02-AUG-1999;	99US-0146388.	PR	28-OCT-1999;	99US-0161920.
PR	03-AUG-1999;	99US-0146389.	PR	28-OCT-1999;	99US-0161992.
PR	04-AUG-1999;	99US-0147038.	PR	28-OCT-1999;	99US-0161993.
PR	04-AUG-1999;	99US-0147204.	PR	29-OCT-1999;	99US-0162142.
PR	05-AUG-1999;	99US-0147302.			
PR	05-AUG-1999;	99US-0147192.			
PR	05-AUG-1999;	99US-0147260.			

Query Match 79.2%; Score 312; DB 21; Length 73;  
Best Local Similarity 80.6%; Pred. No. 2.8e-32;

Matches		58;	Conservative	6;	Mismatches	8;	Indels	0;	Gaps	0;
Qy	1	MIEVVCNDRGKVRVKCNKTDGDLKLLIAAQGTGRNKKVLFKDWVSLGSDY 60								
Db	1	mievvcndrgkvrkcnktdgdlklliaaqtgrnkkvlfkdwvslgskdy 60								
Qy	61	EHDGMNLELY 72								
Db	61	ehdgmglely 72								
RESULT 6										
G27281										
XX	ID	G27281 standard; Protein; 73 AA.								
XX	AC	G27281;								
XX		17-OCT-2000 (first entry)								
XX	DE	Zea mays protein fragment SEQ ID NO: 32055.								
XX	KW	Protein identification; signal transduction pathway; metabolic pathway;								
KW		hybridisation assay; genetic mapping; gene expression control; promoter;								
KW		termination sequence; corn.								
XX		Zea mays subsp. mays.								
XX	OS	EP1033405-A2.								
XX	PN	06-SEP-2000.								
XX	PD	25-FEB-2000; 2000EP-0301439.								
XX	PF	25-FEB-1999; 99US-0121825.								
XX	PR	05-MAR-1999; 99US-0123180.								
PR		09-MAR-1999; 99US-0123548.								
PR		23-MAR-1999; 99US-0125788.								
PR		25-MAR-1999; 99US-0126264.								
PR		29-MAR-1999; 99US-0126785.								
PR		01-APR-1999; 99US-0127462.								
PR		06-APR-1999; 99US-0128234.								
PR		08-APR-1999; 99US-0128714.								
PR		16-APR-1999; 99US-0129845.								
PR		19-APR-1999; 99US-0130077.								
PR		21-APR-1999; 99US-0130449.								
PR		23-APR-1999; 99US-0130510.								
PR		28-APR-1999; 99US-0130891.								
PR		30-APR-1999; 99US-0131449.								
PR		04-MAY-1999; 99US-0132048.								
PR		04-MAY-1999; 99US-0132407.								
PR		05-MAY-1999; 99US-0132484.								
PR		05-MAY-1999; 99US-0132485.								
PR		06-MAY-1999; 99US-0132486.								
PR		06-MAY-1999; 99US-0132487.								
PR		07-MAY-1999; 99US-0132863.								
PR		11-MAY-1999; 99US-0134256.								
PR		14-MAY-1999; 99US-0134218.								
PR		14-MAY-1999; 99US-0134219.								
PR		14-MAY-1999; 99US-0134221.								
PR		14-MAY-1999; 99US-0134370.								
PR		18-MAY-1999; 99US-0134768.								
PR		19-MAY-1999; 99US-0134941.								
PR		20-MAY-1999; 99US-0135124.								
PR		21-MAY-1999; 99US-0135353.								
PR		24-MAY-1999; 99US-0135629.								
PR		25-MAY-1999; 99US-0136021.								
PR		27-MAY-1999; 99US-0136392.								
PR		28-MAY-1999; 99US-0136782.								
PR		01-JUN-1999; 99US-0137222.								
PR		03-JUN-1999; 99US-0137528.								
PR		04-JUN-1999; 99US-0137502.								
PR		07-JUN-1999; 99US-0137724.								
PR		08-JUN-1999; 99US-0138094.								



RESULT 8  
 B36291  
 ID B36291 standard; Protein; 32 AA.  
 XX  
 AC B36291;  
 XX  
 DT 23-FEB-2001 (first entry)  
 XX  
 DE Human beacon (short form) protein sequence.  
 XX  
 KW Israeli sand rat; beacon; obesity; NIDDM; energy balance; diabetes;  
 KW ligand; human.  
 XX  
 OS Homo sapiens.  
 XX  
 WO200064931-A1.  
 XX  
 DT 02-NOV-2000.  
 XX  
 PF 19-APR-2000; 2000WO-AU00342.  
 XX  
 PF 23-APR-1999; 99AU-0009919.  
 PR  
 PR 24-MAR-2000; 2000AU-0006454.  
 XX  
 XX (AUTO-) AUTOGEN PTY LTD.  
 XX  
 XX Collier G, Walder K, Zimmet P;  
 PI  
 XX WPI; 2000-687311/67.  
 DR  
 DR N-PSDB; C81770.  
 XX  
 XX Ligand of beacon protein useful for treating obesity, anorexia, energy  
 PT imbalance, diabetes, metabolic syndrome, dyslipidemia, hypertension and  
 PT insulin resistance -  
 XX  
 XX Disclosure; Fig 1b; 67pp; English.  
 PS  
 XX The present invention is related to the isolation of a ligand known as  
 CC beacon from the Israeli sand rat. Beacon is associated with the  
 CC regulation of energy balance, and the protein, its coding sequence and  
 CC analogues can be used in the treatment of diabetes, obesity, anorexia,  
 CC energy imbalance, metabolic syndrome, dyslipidemia, hypertension and  
 CC insulin resistance. In addition, they can be used in agriculture to  
 CC produce leaner animals.  
 CC  
 CC Sequence 32 AA:  
 CC  
 CC Query Match 37.4%; Score 147.5; DB 21; Length 32;  
 CC Best Local Similarity 93.9%; Pred. No. 6.8e-12;  
 CC Matches 31; Conservative 0; Mismatches 1; Indels 1; Gaps 1;  
 QY 1 MIEVVCNDRLGKRVKVCNTDDTFIGDLKLIAA 33  
 DB 1 mievvcndrlgkvvkvcntddttdtgd-kliaa 32  
 RESULT 9  
 G38562  
 ID G38562 standard; Protein; 477 AA.  
 XX  
 AC G38562;  
 XX  
 DT 18-OCT-2000 (first entry)  
 XX  
 DE Arabidopsis thaliana protein fragment SEQ ID NO: 47590.  
 XX  
 KW Protein identification; signal transduction pathway; metabolic pathway;  
 KW hybridisation assay; genetic mapping; gene expression control; promoter;  
 KW termination sequence.  
 XX

OS Arabidopsis thaliana.  
 XX  
 PN EP1033405-A2.  
 XX  
 PD 06-SEP-2000.  
 XX  
 PF 25-FEB-2000; 2000EP-0301439.  
 XX  
 XX 25-FEB-1999; 99US-0121825.  
 PR 05-MAR-1999; 99US-0123180.  
 PR 09-MAR-1999; 99US-0123548.  
 PR 23-MAR-1999; 99US-0125788.  
 PR 23-MAR-1999; 99US-0126284.  
 PR 29-MAR-1999; 99US-0126785.  
 PR 01-APR-1999; 99US-0127462.  
 PR 06-APR-1999; 99US-0128234.  
 PR 08-APR-1999; 99US-0128714.  
 PR 16-APR-1999; 99US-0129845.  
 PR 19-APR-1999; 99US-0130077.  
 PR 21-APR-1999; 99US-0130449.  
 PR 23-APR-1999; 99US-0130510.  
 PR 28-APR-1999; 99US-0130891.  
 PR 30-APR-1999; 99US-0131449.  
 PR 30-APR-1999; 99US-0132048.  
 PR 04-MAY-1999; 99US-0132407.  
 PR 05-MAY-1999; 99US-0132484.  
 PR 06-MAY-1999; 99US-0132486.  
 PR 06-MAY-1999; 99US-0132487.  
 PR 07-MAY-1999; 99US-0132863.  
 PR 11-MAY-1999; 99US-0134256.  
 PR 14-MAY-1999; 99US-0134218.  
 PR 14-MAY-1999; 99US-0134219.  
 PR 14-MAY-1999; 99US-0134221.  
 PR 14-MAY-1999; 99US-0134370.  
 PR 18-MAY-1999; 99US-0134768.  
 PR 20-MAY-1999; 99US-0134941.  
 PR 21-MAY-1999; 99US-0135124.  
 PR 24-MAY-1999; 99US-0135353.  
 PR 25-MAY-1999; 99US-0135629.  
 PR 27-MAY-1999; 99US-0136021.  
 PR 28-MAY-1999; 99US-0136392.  
 PR 01-JUN-1999; 99US-0136782.  
 PR 03-JUN-1999; 99US-0137222.  
 PR 04-JUN-1999; 99US-0137528.  
 PR 07-JUN-1999; 99US-0137502.  
 PR 08-JUN-1999; 99US-0137724.  
 PR 10-JUN-1999; 99US-0138094.  
 PR 10-JUN-1999; 99US-0138540.  
 PR 14-JUN-1999; 99US-0138847.  
 PR 16-JUN-1999; 99US-0139119.  
 PR 17-JUN-1999; 99US-0139452.  
 PR 18-JUN-1999; 99US-0139453.  
 PR 18-JUN-1999; 99US-0139454.  
 PR 18-JUN-1999; 99US-0139455.  
 PR 18-JUN-1999; 99US-0139456.  
 PR 18-JUN-1999; 99US-0139457.  
 PR 18-JUN-1999; 99US-0139458.  
 PR 18-JUN-1999; 99US-0139459.  
 PR 18-JUN-1999; 99US-0139460.  
 PR 18-JUN-1999; 99US-0139461.  
 PR 18-JUN-1999; 99US-0139462.  
 PR 18-JUN-1999; 99US-0139463.  
 PR 18-JUN-1999; 99US-0139464.  
 PR 21-JUN-1999; 99US-0139763.  
 PR 21-JUN-1999; 99US-0139817.  
 PR 23-JUN-1999; 99US-0139899.  
 PR 23-JUN-1999; 99US-0140353.  
 PR 23-JUN-1999; 99US-0140354.  
 PR 24-JUN-1999; 99US-0140695.  
 PR 28-JUN-1999; 99US-0140823.  
 PR 29-JUN-1999; 99US-0140991.

PR 30-JUN-1999; 99US-0141287.  
PR 01-JUL-1999; 99US-0141842.  
PR 01-JUL-1999; 99US-0142154.  
PR 02-JUL-1999; 99US-0142055.  
PR 06-JUL-1999; 99US-0142390.  
PR 08-JUL-1999; 99US-0142803.  
PR 09-JUL-1999; 99US-0142920.  
PR 12-JUL-1999; 99US-0142977.  
PR 13-JUL-1999; 99US-0143542.  
PR 14-JUL-1999; 99US-0143624.  
PR 15-JUL-1999; 99US-0144005.  
PR 16-JUL-1999; 99US-0144085.  
PR 16-JUL-1999; 99US-0144086.  
PR 19-JUL-1999; 99US-0144325.  
PR 19-JUL-1999; 99US-0144331.  
PR 19-JUL-1999; 99US-0144332.  
PR 19-JUL-1999; 99US-0144333.  
PR 19-JUL-1999; 99US-0144334.  
PR 19-JUL-1999; 99US-0144335.  
PR 20-JUL-1999; 99US-0144352.  
PR 20-JUL-1999; 99US-0144632.  
PR 20-JUL-1999; 99US-0144884.  
PR 21-JUL-1999; 99US-0144814.  
PR 21-JUL-1999; 99US-0145086.  
PR 21-JUL-1999; 99US-0145088.  
PR 22-JUL-1999; 99US-0145085.  
PR 22-JUL-1999; 99US-0145087.  
PR 22-JUL-1999; 99US-0145089.  
PR 22-JUL-1999; 99US-0145192.  
PR 23-JUL-1999; 99US-0145145.  
PR 23-JUL-1999; 99US-0145218.  
PR 23-JUL-1999; 99US-0145224.  
PR 26-JUL-1999; 99US-0145276.  
PR 27-JUL-1999; 99US-0145913.  
PR 27-JUL-1999; 99US-0145918.  
PR 27-JUL-1999; 99US-0145919.  
PR 28-JUL-1999; 99US-0145951.  
PR 02-AUG-1999; 99US-0146386.  
PR 02-AUG-1999; 99US-0146388.  
PR 02-AUG-1999; 99US-0146389.  
PR 02-AUG-1999; 99US-0146399.  
PR 03-AUG-1999; 99US-0147038.  
PR 04-AUG-1999; 99US-0147204.  
PR 04-AUG-1999; 99US-0147302.  
PR 05-AUG-1999; 99US-0147192.  
PR 05-AUG-1999; 99US-0147260.  
PR 06-AUG-1999; 99US-0147303.  
PR 06-AUG-1999; 99US-0147416.  
PR 09-AUG-1999; 99US-0147493.  
PR 09-AUG-1999; 99US-0147935.  
PR 10-AUG-1999; 99US-0148171.  
PR 11-AUG-1999; 99US-0148319.  
PR 12-AUG-1999; 99US-0148341.  
PR 13-AUG-1999; 99US-0148565.  
PR 13-AUG-1999; 99US-0148684.  
PR 16-AUG-1999; 99US-0149368.  
PR 17-AUG-1999; 99US-0149175.  
PR 18-AUG-1999; 99US-0149426.  
PR 20-AUG-1999; 99US-0149722.  
PR 20-AUG-1999; 99US-0149723.  
PR 20-AUG-1999; 99US-0149929.  
PR 23-AUG-1999; 99US-0149902.  
PR 23-AUG-1999; 99US-0149930.  
PR 25-AUG-1999; 99US-0150566.  
PR 26-AUG-1999; 99US-0150884.  
PR 27-AUG-1999; 99US-0151065.  
PR 27-AUG-1999; 99US-0151066.  
PR 27-AUG-1999; 99US-0151080.  
PR 30-AUG-1999; 99US-0151303.  
PR 31-AUG-1999; 99US-0151438.  
PR 01-SEP-1999; 99US-0151930.  
PR 07-SEP-1999; 99US-0152363.  
PR 10-SEP-1999; 99US-0153070.  
PR 13-SEP-1999; 99US-0153758.

PR 15-SEP-1999; 99US-0154018.  
PR 16-SEP-1999; 99US-0154039.  
PR 20-SEP-1999; 99US-0154779.  
PR 22-SEP-1999; 99US-0155139.  
PR 23-SEP-1999; 99US-0155486.  
PR 24-SEP-1999; 99US-0155659.  
PR 28-SEP-1999; 99US-0156458.  
PR 29-SEP-1999; 99US-0156596.  
PR 04-OCT-1999; 99US-0157117.  
PR 05-OCT-1999; 99US-0157753.  
PR 06-OCT-1999; 99US-0157865.  
PR 07-OCT-1999; 99US-0158029.  
PR 08-OCT-1999; 99US-0158232.  
PR 12-OCT-1999; 99US-0158369.  
PR 13-OCT-1999; 99US-0159293.  
PR 13-OCT-1999; 99US-0159294.  
PR 13-OCT-1999; 99US-0159295.  
PR 14-OCT-1999; 99US-0159329.  
PR 14-OCT-1999; 99US-0159330.  
PR 14-OCT-1999; 99US-0159331.  
PR 14-OCT-1999; 99US-0159637.  
PR 14-OCT-1999; 99US-0159638.  
PR 18-OCT-1999; 99US-0159584.  
PR 21-OCT-1999; 99US-0160741.  
PR 21-OCT-1999; 99US-0160767.  
PR 21-OCT-1999; 99US-0160768.  
PR 21-OCT-1999; 99US-0160770.  
PR 21-OCT-1999; 99US-0160814.  
PR 21-OCT-1999; 99US-0160815.  
PR 22-OCT-1999; 99US-0160980.  
PR 22-OCT-1999; 99US-0160981.  
PR 22-OCT-1999; 99US-0160989.  
PR 25-OCT-1999; 99US-0161404.  
PR 25-OCT-1999; 99US-0161405.  
PR 25-OCT-1999; 99US-0161406.  
PR 26-OCT-1999; 99US-0161359.  
PR 26-OCT-1999; 99US-0161360.  
PR 26-OCT-1999; 99US-0161361.  
PR 28-OCT-1999; 99US-0161920.  
PR 28-OCT-1999; 99US-0161922.  
PR 28-OCT-1999; 99US-0161993.  
PR 29-OCT-1999; 99US-0162142.

Query Match 20.3%; Score 80; DB 21; Length 477;  
Best Local Similarity 31.7%; Pred. No. 0.073;

Matches 19; Conservative 16; Mismatches 25; Indels 0; Gaps 0;

Qy 11 GKRVKCNFTDTIGDLKLIAAQCTRNKIVLKKWYIFKDHVSLGDYEIHGNNLEL 70  
Db 407 gktillevessdtianvkekigkpdqgmliiffgqgledgvtlgydihkkslyl 466

RESULT 10

G38561  
ID G38561 standard; Protein; 553 AA.

XX G38561;

XX AC  
XX XX  
DT 18-OCT-2000 (first entry)

XX DE Arabidopsis thaliana protein fragment SEQ ID NO: 47589.

XX KW Protein identification; signal transduction pathway; metabolic pathway;  
XX KW hybridisation assay; genetic mapping; gene expression control; promoter;  
XX KW termination sequence.

XX OS Arabidopsis thaliana.

XX PN EP1033405-A2.

XX PD 06-SEP-2000.

XX PF 25-FEB-2000; 2000EP-0301439.

XX PR 25-FEB-1999; 99US-0121825.  
PR 05-MAR-1999; 99US-0123180.  
PR 09-MAR-1999; 99US-0123548.  
PR 23-MAR-1999; 99US-0125788.  
PR 25-MAR-1999; 99US-0126264.  
PR 29-MAR-1999; 99US-0126785.  
PR 01-APR-1999; 99US-0127462.  
PR 06-APR-1999; 99US-0128234.  
PR 08-APR-1999; 99US-0128714.  
PR 16-APR-1999; 99US-0129845.  
PR 19-APR-1999; 99US-0130077.  
PR 21-APR-1999; 99US-0130449.  
PR 23-APR-1999; 99US-0130510.  
PR 28-APR-1999; 99US-0130891.  
PR 30-APR-1999; 99US-0131419.  
PR 30-APR-1999; 99US-0132048.  
PR 30-APR-1999; 99US-0132407.  
PR 04-MAY-1999; 99US-0132484.  
PR 05-MAY-1999; 99US-0132485.  
PR 06-MAY-1999; 99US-0132486.  
PR 06-MAY-1999; 99US-0132487.  
PR 07-MAY-1999; 99US-0132863.  
PR 11-MAY-1999; 99US-0134256.  
PR 14-MAY-1999; 99US-0134218.  
PR 14-MAY-1999; 99US-0134219.  
PR 14-MAY-1999; 99US-0134221.  
PR 14-MAY-1999; 99US-0134370.  
PR 18-MAY-1999; 99US-0134768.  
PR 19-MAY-1999; 99US-0134941.  
PR 20-MAY-1999; 99US-0135124.  
PR 21-MAY-1999; 99US-0135353.  
PR 24-MAY-1999; 99US-0135629.  
PR 25-MAY-1999; 99US-0136021.  
PR 27-MAY-1999; 99US-0136392.  
PR 28-MAY-1999; 99US-0136782.  
PR 01-JUN-1999; 99US-0137222.  
PR 03-JUN-1999; 99US-0137528.  
PR 04-JUN-1999; 99US-0137502.  
PR 07-JUN-1999; 99US-0137724.  
PR 08-JUN-1999; 99US-0138094.  
PR 10-JUN-1999; 99US-0138540.  
PR 10-JUN-1999; 99US-0138847.  
PR 14-JUN-1999; 99US-0139119.  
PR 16-JUN-1999; 99US-0139452.  
PR 16-JUN-1999; 99US-0139453.  
PR 17-JUN-1999; 99US-0139492.  
PR 18-JUN-1999; 99US-0139454.  
PR 18-JUN-1999; 99US-0139455.  
PR 18-JUN-1999; 99US-0139456.  
PR 18-JUN-1999; 99US-0139457.  
PR 18-JUN-1999; 99US-0139458.  
PR 18-JUN-1999; 99US-0139459.  
PR 18-JUN-1999; 99US-0139460.  
PR 18-JUN-1999; 99US-0139461.  
PR 18-JUN-1999; 99US-0139462.  
PR 18-JUN-1999; 99US-0139463.  
PR 18-JUN-1999; 99US-0139750.  
PR 18-JUN-1999; 99US-0139753.  
PR 21-JUN-1999; 99US-0139817.  
PR 22-JUN-1999; 99US-0139899.  
PR 23-JUN-1999; 99US-0140353.  
PR 23-JUN-1999; 99US-0140354.  
PR 24-JUN-1999; 99US-0140699.  
PR 28-JUN-1999; 99US-0140823.  
PR 29-JUN-1999; 99US-0140991.  
PR 30-JUN-1999; 99US-0141287.  
PR 01-JUL-1999; 99US-0141842.  
PR 01-JUL-1999; 99US-0142154.  
PR 02-JUL-1999; 99US-0142059.  
PR 06-JUL-1999; 99US-0142390.  
PR 08-JUL-1999; 99US-0142803.  
PR 09-JUL-1999; 99US-0142920.  
PR 12-JUL-1999; 99US-0142977.  
PR 13-JUL-1999; 99US-0143542.  
PR 14-JUL-1999; 99US-0143624.  
PR 15-JUL-1999; 99US-0144005.  
PR 16-JUL-1999; 99US-0144085.  
PR 16-JUL-1999; 99US-0144086.  
PR 19-JUL-1999; 99US-0144325.  
PR 19-JUL-1999; 99US-0144331.  
PR 19-JUL-1999; 99US-0144332.  
PR 19-JUL-1999; 99US-0144333.  
PR 19-JUL-1999; 99US-0144334.  
PR 19-JUL-1999; 99US-0144335.  
PR 20-JUL-1999; 99US-0144352.  
PR 20-JUL-1999; 99US-0144632.  
PR 20-JUL-1999; 99US-0144684.  
PR 21-JUL-1999; 99US-0144814.  
PR 21-JUL-1999; 99US-0145086.  
PR 21-JUL-1999; 99US-0145088.  
PR 22-JUL-1999; 99US-0145085.  
PR 22-JUL-1999; 99US-0145087.  
PR 22-JUL-1999; 99US-0145089.  
PR 22-JUL-1999; 99US-0145192.  
PR 23-JUL-1999; 99US-0145145.  
PR 23-JUL-1999; 99US-0145218.  
PR 23-JUL-1999; 99US-0145224.  
PR 26-JUL-1999; 99US-0145276.  
PR 27-JUL-1999; 99US-0145913.  
PR 27-JUL-1999; 99US-0145918.  
PR 27-JUL-1999; 99US-0145919.  
PR 28-JUL-1999; 99US-0145951.  
PR 02-AUG-1999; 99US-0146386.  
PR 02-AUG-1999; 99US-0146388.  
PR 02-AUG-1999; 99US-0146389.  
PR 03-AUG-1999; 99US-0147038.  
PR 04-AUG-1999; 99US-0147204.  
PR 04-AUG-1999; 99US-0147302.  
PR 05-AUG-1999; 99US-0147192.  
PR 05-AUG-1999; 99US-0147260.  
PR 06-AUG-1999; 99US-0147303.  
PR 06-AUG-1999; 99US-0147416.  
PR 09-AUG-1999; 99US-0147493.  
PR 09-AUG-1999; 99US-0147935.  
PR 10-AUG-1999; 99US-0148171.  
PR 11-AUG-1999; 99US-0148319.  
PR 12-AUG-1999; 99US-0148341.  
PR 13-AUG-1999; 99US-0148565.  
PR 13-AUG-1999; 99US-0148684.  
PR 16-AUG-1999; 99US-0149368.  
PR 17-AUG-1999; 99US-0149175.  
PR 18-AUG-1999; 99US-0149426.  
PR 20-AUG-1999; 99US-0149722.  
PR 20-AUG-1999; 99US-0149723.  
PR 20-AUG-1999; 99US-0149929.  
PR 23-AUG-1999; 99US-0149902.  
PR 23-AUG-1999; 99US-0149930.  
PR 25-AUG-1999; 99US-0150566.  
PR 26-AUG-1999; 99US-0150884.  
PR 27-AUG-1999; 99US-0151065.  
PR 27-AUG-1999; 99US-0151066.  
PR 30-AUG-1999; 99US-0151080.  
PR 30-AUG-1999; 99US-0151303.  
PR 31-AUG-1999; 99US-0151438.  
PR 01-SEP-1999; 99US-0151930.  
PR 07-SEP-1999; 99US-0152363.  
PR 10-SEP-1999; 99US-0153070.  
PR 13-SEP-1999; 99US-0153758.  
PR 15-SEP-1999; 99US-0154018.  
PR 16-SEP-1999; 99US-0154039.  
PR 20-SEP-1999; 99US-0154779.  
PR 22-SEP-1999; 99US-0155139.  
PR 23-SEP-1999; 99US-0155486.  
PR 24-SEP-1999; 99US-0155659.  
PR 28-SEP-1999; 99US-0156458.









XX Protein identification; signal transduction pathway; metabolic pathway;  
KW hybridisation assay; genetic mapping; gene expression control; promoter;  
KW termination sequence.  
XX Arabidopsis thaliana.  
XX EP1033405-A2.  
XX 06-SEP-2000.  
XX 25-FEB-2000; 2000EP-0301439.  
XX 25-FEB-1999; 99US-0121825.  
PR 05-MAR-1999; 99US-0123180.  
PR 09-MAR-1999; 99US-0123548.  
PR 23-MAR-1999; 99US-0125788.  
PR 25-MAR-1999; 99US-0126264.  
PR 29-MAR-1999; 99US-0126785.  
PR 01-APR-1999; 99US-0127462.  
PR 06-APR-1999; 99US-0128234.  
PR 08-APR-1999; 99US-0128714.  
PR 16-APR-1999; 99US-0129845.  
PR 19-APR-1999; 99US-0130077.  
PR 21-APR-1999; 99US-0130449.  
PR 23-APR-1999; 99US-0130510.  
PR 23-APR-1999; 99US-0130891.  
PR 28-APR-1999; 99US-0131449.  
PR 30-APR-1999; 99US-0132048.  
PR 30-APR-1999; 99US-0132407.  
PR 04-MAY-1999; 99US-0132484.  
PR 05-MAY-1999; 99US-0132485.  
PR 06-MAY-1999; 99US-0132486.  
PR 06-MAY-1999; 99US-0132487.  
PR 07-MAY-1999; 99US-0132863.  
PR 11-MAY-1999; 99US-0134256.  
PR 14-MAY-1999; 99US-0134218.  
PR 14-MAY-1999; 99US-0134219.  
PR 14-MAY-1999; 99US-0134221.  
PR 14-MAY-1999; 99US-0134370.  
PR 18-MAY-1999; 99US-0134768.  
PR 19-MAY-1999; 99US-0134941.  
PR 20-MAY-1999; 99US-0135124.  
PR 21-MAY-1999; 99US-0135353.  
PR 24-MAY-1999; 99US-0135629.  
PR 25-MAY-1999; 99US-0136021.  
PR 27-MAY-1999; 99US-0136392.  
PR 28-MAY-1999; 99US-0136782.  
PR 01-JUN-1999; 99US-0137222.  
PR 03-JUN-1999; 99US-0137528.  
PR 04-JUN-1999; 99US-0137502.  
PR 07-JUN-1999; 99US-0137724.  
PR 08-JUN-1999; 99US-0138094.  
PR 10-JUN-1999; 99US-0138540.  
PR 10-JUN-1999; 99US-0138847.  
PR 14-JUN-1999; 99US-0139119.  
PR 16-JUN-1999; 99US-0139452.  
PR 16-JUN-1999; 99US-0139453.  
PR 17-JUN-1999; 99US-0139492.  
PR 18-JUN-1999; 99US-0139454.  
PR 18-JUN-1999; 99US-0139455.  
PR 18-JUN-1999; 99US-0139456.  
PR 18-JUN-1999; 99US-0139457.  
PR 18-JUN-1999; 99US-0139458.  
PR 18-JUN-1999; 99US-0139459.  
PR 18-JUN-1999; 99US-0139460.  
PR 18-JUN-1999; 99US-0139461.  
PR 18-JUN-1999; 99US-0139462.  
PR 18-JUN-1999; 99US-0139463.  
PR 18-JUN-1999; 99US-0139750.  
PR 18-JUN-1999; 99US-0139763.  
PR 21-JUN-1999; 99US-0139817.  
PR 22-JUN-1999; 99US-0139899.  
PR 23-JUN-1999; 99US-0140353.  
PR 23-JUN-1999; 99US-0140354.  
PR 24-JUN-1999; 99US-0140695.  
PR 28-JUN-1999; 99US-0140823.  
PR 29-JUN-1999; 99US-0140991.  
PR 30-JUN-1999; 99US-0141287.  
PR 01-JUL-1999; 99US-0141842.  
PR 01-JUL-1999; 99US-0142154.  
PR 02-JUL-1999; 99US-0142055.  
PR 06-JUL-1999; 99US-0142390.  
PR 08-JUL-1999; 99US-0142803.  
PR 09-JUL-1999; 99US-0142920.  
PR 12-JUL-1999; 99US-0142977.  
PR 13-JUL-1999; 99US-0143542.  
PR 14-JUL-1999; 99US-0143624.  
PR 15-JUL-1999; 99US-0144005.  
PR 16-JUL-1999; 99US-0144085.  
PR 16-JUL-1999; 99US-0144086.  
PR 19-JUL-1999; 99US-0144325.  
PR 19-JUL-1999; 99US-0144331.  
PR 19-JUL-1999; 99US-0144332.  
PR 19-JUL-1999; 99US-0144333.  
PR 19-JUL-1999; 99US-0144334.  
PR 19-JUL-1999; 99US-0144335.  
PR 20-JUL-1999; 99US-0144352.  
PR 20-JUL-1999; 99US-0144632.  
PR 20-JUL-1999; 99US-0144884.  
PR 21-JUL-1999; 99US-0144814.  
PR 21-JUL-1999; 99US-0145086.  
PR 21-JUL-1999; 99US-0145088.  
PR 22-JUL-1999; 99US-0145085.  
PR 22-JUL-1999; 99US-0145087.  
PR 22-JUL-1999; 99US-0145089.  
PR 22-JUL-1999; 99US-0145192.  
PR 23-JUL-1999; 99US-0145145.  
PR 23-JUL-1999; 99US-0145218.  
PR 23-JUL-1999; 99US-0145224.  
PR 26-JUL-1999; 99US-0145276.  
PR 27-JUL-1999; 99US-0145913.  
PR 27-JUL-1999; 99US-0145918.  
PR 27-JUL-1999; 99US-0145919.  
PR 28-JUL-1999; 99US-0145951.  
PR 02-AUG-1999; 99US-0146386.  
PR 02-AUG-1999; 99US-0146388.  
PR 02-AUG-1999; 99US-0146389.  
PR 03-AUG-1999; 99US-0147038.  
PR 04-AUG-1999; 99US-0147204.  
PR 04-AUG-1999; 99US-0147302.  
PR 05-AUG-1999; 99US-0147192.  
PR 05-AUG-1999; 99US-0147260.  
PR 06-AUG-1999; 99US-0147303.  
PR 06-AUG-1999; 99US-0147416.  
PR 09-AUG-1999; 99US-0147493.  
PR 09-AUG-1999; 99US-0147935.  
PR 10-AUG-1999; 99US-0148171.  
PR 11-AUG-1999; 99US-0148319.  
PR 12-AUG-1999; 99US-0148341.  
PR 13-AUG-1999; 99US-0148565.  
PR 13-AUG-1999; 99US-0148684.  
PR 16-AUG-1999; 99US-0149368.  
PR 17-AUG-1999; 99US-0149175.  
PR 18-AUG-1999; 99US-0149426.  
PR 20-AUG-1999; 99US-0149722.  
PR 20-AUG-1999; 99US-0149723.  
PR 20-AUG-1999; 99US-0149929.  
PR 23-AUG-1999; 99US-0149902.  
PR 23-AUG-1999; 99US-0149930.  
PR 25-AUG-1999; 99US-0150566.  
PR 26-AUG-1999; 99US-0150884.  
PR 27-AUG-1999; 99US-0151065.  
PR 27-AUG-1999; 99US-0151066.  
PR 27-AUG-1999; 99US-0151080.  
PR 30-AUG-1999; 99US-0151303.

PR	31-AUG-1999;	99US-0151438.	PN	EPI033405-A2.
PR	01-SEP-1999;	99US-0151930.	XX	
PR	07-SEP-1999;	99US-0152363.	PD	06-SEP-2000.
PR	10-SEP-1999;	99US-0153070.	XX	
PR	13-SEP-1999;	99US-0153758.	PF	25-FEB-2000; 2000EP-0301439.
PR	15-SEP-1999;	99US-0154018.	XX	
PR	16-SEP-1999;	99US-0154039.	XX	25-FEB-1999; 99US-0121825.
PR	20-SEP-1999;	99US-0154779.	PR	05-MAR-1999; 99US-0123180.
PR	22-SEP-1999;	99US-0155139.	PR	09-MAR-1999; 99US-0123548.
PR	23-SEP-1999;	99US-0155186.	PR	23-MAR-1999; 99US-0125788.
PR	24-SEP-1999;	99US-0155659.	PR	25-MAR-1999; 99US-0126264.
PR	28-SEP-1999;	99US-0156458.	PR	29-MAR-1999; 99US-0126785.
PR	29-SEP-1999;	99US-0156596.	PR	01-APR-1999; 99US-0127482.
PR	04-OCT-1999;	99US-0157117.	PR	06-APR-1999; 99US-0128234.
PR	05-OCT-1999;	99US-0157753.	PR	08-APR-1999; 99US-0128714.
PR	06-OCT-1999;	99US-0157865.	PR	16-APR-1999; 99US-0129845.
PR	07-OCT-1999;	99US-0158029.	PR	19-APR-1999; 99US-0130077.
PR	08-OCT-1999;	99US-0158232.	PR	21-APR-1999; 99US-0130449.
PR	12-OCT-1999;	99US-0158369.	PR	23-APR-1999; 99US-0130510.
PR	13-OCT-1999;	99US-0159293.	PR	23-APR-1999; 99US-0130891.
PR	13-OCT-1999;	99US-0159294.	PR	28-APR-1999; 99US-0131449.
PR	13-OCT-1999;	99US-0159295.	PR	30-APR-1999; 99US-0132048.
PR	14-OCT-1999;	99US-0159329.	PR	30-APR-1999; 99US-0132407.
PR	14-OCT-1999;	99US-0159330.	PR	04-MAY-1999; 99US-0132484.
PR	14-OCT-1999;	99US-0159331.	PR	05-MAY-1999; 99US-0132485.
PR	14-OCT-1999;	99US-0159637.	PR	06-MAY-1999; 99US-0132486.
PR	14-OCT-1999;	99US-0159638.	PR	06-MAY-1999; 99US-0132487.
PR	18-OCT-1999;	99US-0159584.	PR	07-MAY-1999; 99US-0132863.
PR	21-OCT-1999;	99US-0160741.	PR	11-MAY-1999; 99US-0134256.
PR	21-OCT-1999;	99US-0160767.	PR	14-MAY-1999; 99US-0134218.
PR	21-OCT-1999;	99US-0160770.	PR	14-MAY-1999; 99US-0134219.
PR	21-OCT-1999;	99US-0160814.	PR	14-MAY-1999; 99US-0134221.
PR	21-OCT-1999;	99US-0160815.	PR	14-MAY-1999; 99US-0134370.
PR	22-OCT-1999;	99US-0160980.	PR	18-MAY-1999; 99US-0134768.
PR	22-OCT-1999;	99US-0160981.	PR	19-MAY-1999; 99US-0134941.
PR	22-OCT-1999;	99US-0160989.	PR	20-MAY-1999; 99US-0135124.
PR	25-OCT-1999;	99US-0161404.	PR	21-MAY-1999; 99US-0135353.
PR	25-OCT-1999;	99US-0161405.	PR	24-MAY-1999; 99US-0135629.
PR	25-OCT-1999;	99US-0161406.	PR	25-MAY-1999; 99US-0136021.
PR	26-OCT-1999;	99US-0161359.	PR	27-MAY-1999; 99US-0136392.
PR	26-OCT-1999;	99US-0161360.	PR	28-MAY-1999; 99US-0136782.
PR	26-OCT-1999;	99US-0161361.	PR	01-JUN-1999; 99US-0137222.
PR	28-OCT-1999;	99US-0161920.	PR	03-JUN-1999; 99US-0137528.
PR	28-OCT-1999;	99US-0161992.	PR	04-JUN-1999; 99US-0137502.
PR	28-OCT-1999;	99US-0161993.	PR	07-JUN-1999; 99US-0137724.
PR	28-OCT-1999;	99US-0161993.	PR	08-JUN-1999; 99US-0138094.
PR	29-OCT-1999;	99US-0162142.	PR	10-JUN-1999; 99US-0138540.
PR			PR	10-JUN-1999; 99US-0138847.
PR			PR	14-JUN-1999; 99US-0139119.
PR			PR	16-JUN-1999; 99US-0139452.
PR			PR	16-JUN-1999; 99US-0139453.
PR			PR	17-JUN-1999; 99US-0139492.
PR			PR	18-JUN-1999; 99US-0139454.
PR			PR	18-JUN-1999; 99US-0139455.
PR			PR	18-JUN-1999; 99US-0139456.
PR			PR	18-JUN-1999; 99US-0139457.
PR			PR	18-JUN-1999; 99US-0139458.
PR			PR	18-JUN-1999; 99US-0139459.
PR			PR	18-JUN-1999; 99US-0139460.
PR			PR	18-JUN-1999; 99US-0139461.
PR			PR	18-JUN-1999; 99US-0139462.
PR			PR	18-JUN-1999; 99US-0139463.
PR			PR	18-JUN-1999; 99US-0139750.
PR			PR	18-JUN-1999; 99US-0139763.
PR			PR	21-JUN-1999; 99US-0139817.
PR			PR	22-JUN-1999; 99US-0139899.
PR			PR	23-JUN-1999; 99US-0140353.
PR			PR	23-JUN-1999; 99US-0140354.
PR			PR	24-JUN-1999; 99US-0140695.
PR			PR	24-JUN-1999; 99US-0140823.
PR			PR	29-JUN-1999; 99US-0140991.
OS	Arabidopsis thaliana.		PR	30-JUN-1999; 99US-0141287.
XX			PR	01-JUL-1999; 99US-0141842.

Query Match 15.28; Score 60; DB 21; Length 295;

Best Local Similarity 30.8%; Pred. No. 14;

Matches 16; Conservative 9; Mismatches 13; Indels 14; Gaps 2;

11 GKVRVK-----CWTDDTIGDLKLLIAAQGTFRWKNKIVLKKWYTFKD 53

37 girvrkygsvhheinsngsfelgkllsgatgvhqdmgq-----lykd 83

RESULT 15

G23779

ID G23779 standard; Protein; 296 AA.

XX AC

XX AC

XX AC

XX AC

XX AC

XX AC

XX AC

XX AC

XX AC

XX AC

XX AC

XX AC

XX AC

XX AC

XX AC

XX AC

XX AC

XX AC

XX AC

XX AC

XX AC

Arabidopsis thaliana protein fragment SEQ ID NO: 27212.

Protein identification; signal transduction pathway; metabolic pathway; hybridisation assay; genetic mapping; gene expression control; promoter; termination sequence.

Arabidopsis thaliana.



GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: June 13, 2001, 20:25:33 ; Search time 60.15 Seconds  
(without alignments)  
142.247 Million cell updates/sec

Title: US-09-331-930A-2  
Perfect score: 394  
Sequence: 1 MIEVQNDRLGKKVRKCNTP.....HVSIGDVEIHGMNLELYQ 73

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Search: 374700 seqs, 117207915 residues

Total number of hits satisfying chosen parameters: 374700

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

SPTREMBL15:

- 1: sp\_archaea:\*
- 2: sp\_bacteria:\*
- 3: sp\_fungi:\*
- 4: sp\_human:\*
- 5: sp\_invertebrate:\*
- 6: sp\_mammal:\*
- 7: sp\_mhc:\*
- 8: sp\_organalle:\*
- 9: sp\_phase:\*
- 10: sp\_plant:\*
- 11: sp\_rodent:\*
- 12: sp\_unclassified:\*
- 13: sp\_vertebrate:\*
- 14: sp\_virus:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	339	86.0	73	5 Q9V998	Q9V998 drosophila
2	331	84.0	73	5 P91302	P91302 caenorhabdi
3	306	77.7	73	10 Q9M1U1	Q9M1U1 arabidopsis
4	301	76.4	73	3 Q94650	Q94650 schizosacch
5	89	22.6	115	14 Q90461	Q90461 human immun
6	88.5	22.5	114	14 Q9WNO6	Q9WNO6 human immun
7	88	22.3	203	14 Q9JBC9	Q9JBC9 human immun
8	87	22.1	115	14 Q71412	Q71412 human immun
9	87	22.1	215	14 Q89553	Q89553 human immun
10	86	21.8	115	14 Q9WNN3	Q9WNN3 human immun
11	86	21.8	164	14 Q41876	Q41876 human immun
12	85.5	21.7	114	14 Q71495	Q71495 human immun
13	85.5	21.7	114	14 Q90467	Q90467 human immun
14	85.5	21.7	176	14 Q9QCN8	Q9QCN8 human immun
15	85.5	21.7	209	14 Q75118	Q75118 human immun
16	85.5	21.7	417	14 Q9QCN7	Q9QCN7 human immun
17	85.5	21.7	857	14 Q9QML6	Q9QML6 human immun
18	85	21.6	115	14 Q70167	Q70167 human immun
19	85	21.6	115	14 Q70175	Q70175 human immun

20	85	21.6	115	14	Q74431	human immun
21	85	21.6	115	14	Q90411	human immun
22	85	21.6	115	14	Q9YTW7	human immun
23	85	21.6	119	14	Q77156	human immun
24	85	21.6	202	14	Q9JBC1	human immun
25	85	21.6	203	14	Q9JBB7	human immun
26	84.5	21.4	114	14	Q71437	human immun
27	84.5	21.4	114	14	P88770	human immun
28	84.5	21.4	159	14	Q9WBN6	human immun
29	84	21.3	115	14	Q9WNP7	human immun
30	84	21.3	212	14	Q9IVU1	human immun
31	83.5	21.2	101	14	Q9QSW3	human immun
32	83.5	21.2	163	14	Q41868	human immun
33	83.5	21.2	163	14	Q41877	human immun
34	83.5	21.2	164	14	O56899	human immun
35	83.5	21.2	455	14	Q91ZE4	human immun
36	83.5	21.2	543	14	Q9YUY7	human immun
37	83	21.1	175	14	P88226	human immun
38	83	21.1	200	14	Q9JBB2	human immun
39	83	21.1	200	14	Q9JBA7	human immun
40	83	21.1	201	14	Q9JBC2	human immun
41	83	21.1	201	14	Q9JBA4	human immun
42	83	21.1	201	14	Q9JBB1	human immun
43	83	21.1	201	14	Q9JBA9	human immun
44	83	21.1	201	14	Q9JBA6	human immun
45	83	21.1	201	14	Q9JBA5	human immun

## ALIGNMENTS

RESULT	1
Q9V998	PRELIMINARY; PRT; 73 AA.
AC	Q9V998; PRELIMINARY; PRT; 73 AA.
DT	01-MAY-2000 (TREMBLrel. 13, Created)
DT	01-MAY-2000 (TREMBLrel. 13, Last sequence update)
DE	01-JUN-2000 (TREMBLrel. 14, Last annotation update)
DE	CG3450 PROTEIN.
GN	CG3450.
OS	Drosophila melanogaster (Fruit fly).
OC	Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC	Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC	Ephydroidea; Drosophilidae; Drosophila.
OX	NCBI_TaxID=7227;
RN	[1]
RP	SEQUENCE FROM N.A.
RX	STRAIN=BERKELEY;
RX	MEDLINE=20196006; PubMed=10731132;
RA	Adams M.D., Celisner S.E., Holt R.A., Evans C.A., Gocayne J.D.,
RA	Ananatiades P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,
RA	George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
RA	Sutton G.G., Wortman J.R., Vandeil M.D., Zhang Q., Chen L.X.,
RA	Brandon R.C., Rogers Y.-H.C., Blazej R.G., Champagne M., Pfeiffer B.D.,
RA	Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,
RA	Abril J.F., Agbayani A., An H.-J., Andrews-Pfannkuch C., Baldwin D.,
RA	Baliew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
RA	Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,
RA	Borkova D., Botchan M.R., Bouck J., Brokstein P., Brottier P.,
RA	Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
RA	Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
RA	de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
RA	Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
RA	Durbin K.J., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W.,
RA	Foster C., Gabriellian A.E., Garg N.S., Gelbart W.M., Glasser K.,
RA	Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
RA	Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,
RA	Hoslin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,
RA	Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
RA	Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
RA	Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
RA	Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,
RA	Merkulov G., Milshina N.V., Mobarri C., Morris J., Moshrefi A.,





```
OX NCBI_TaxID=4896;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=972;
RA Oliver K., Harris D., Wood V., Rajandream M.A., Barrell B.G.;
RL Submitted (MAR-1999) to the EMBL/GenBank/DBJ databases.
CC -1- SUBCELLULAR LOCATION: NUCLEAR (BY SIMILARITY).
CC -1- SIMILARITY: CONTAINS A UBQUITIN-LIKE DOMAIN.
DR EMBL; AL049190; CAB39137.1; -.
DR INTERPRO; IPR000626; -.
DR PROSITE; PS50053; UBQUITIN_2; 1.
KW Hypothetical protein; Nuclear protein; Multigene family.
FT DOMAIN 2 73 UBQUITIN-LIKE.
SQ SEQUENCE 73 AA; 8435 MW; 894BED5601FE365 CRC64;

Query Match 76.4%; Score 301; DB 3; Length 73;
Best Local Similarity 73.6%; Pred. No. 2.5e-28;
Matches 53; Conservative 12; Mismatches 7; Indels 0; Gaps 0;

QY 1 MIEVWNCNDRLGKRVKVCNTDITGDKLKLIAAQTGRWNKIVLKWKYTIKFDHVSIGDY 60
Db 1 MIEVLCNDRLGKRVKVCNPDITGDKLKLIAAQTGRWNKIVLKWKYTIKFDHVSIGDY 60
QY 61 EIHGGMNLELY 72
Db 61 EIHGGMNLELY 72

RESULT 5
O90461 PRELIMINARY; PRT; 115 AA.
AC O90461;
DT 01-NOV-1998 (TREMBLrel. 08, Created)
DT 01-NOV-1998 (TREMBLrel. 08, Last sequence update)
DT 01-MAY-2000 (TREMBLrel. 13, Last annotation update)
DE GP120 (FRAGMENT).
GN ENV.
OS Human immunodeficiency virus type 1.
OC Viruses; Retroid viruses; Retroviridae; Lentivirus.
OX NCBI_TaxID=11676;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=THSC62;
RA Subbarao S., Vanichseni S., Kitayaporn D., Hart C., Raktham S.,
Young N., Srisuwanalai L., Luo C.C., Mastro T.D.;
"Genetic Characterization of Incident HIV-1 Subtype B and E Strains
from Injecting Drug Users (IDUs) in Bangkok, Thailand.";
RL Submitted (AUG-1998) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF081754; AAC34012.1; -.
DR INTERPRO; IPR000777; -.
DR PFAM; PF00516; GP120; 1.
FT NON_TER 1
FT NON_TER 115
SQ SEQUENCE 115 AA; 13117 MW; 2E15C4A6A3405CD1 CRC64;

Query Match 22.6%; Score 89; DB 14; Length 115;
Best Local Similarity 29.5%; Pred. No. 0.0039;
Matches 28; Conservative 9; Mismatches 26; Indels 32; Gaps 4;

QY 10 LGKRVKVC-----NTDITGDKLKLIAAQTGRWNKIV-----43
Db 13 LKNSVEINCRPSNNTRTSIRIGPGVFTGTGIDIRKAYCEINGTKWVKLVQVTEK 72

QY 44 LKKWY---TIFKDHVSLGDYEI---HDGMNLELY 72
Db 73 LKKHFNKTIIVPEPPSGDLEITTHFNCRGEFFY 107

RESULT 6
O9WNQ6 PRELIMINARY; PRT; 114 AA.
ID O9WNQ6
```

```
AC O9WNQ6;
DT 01-NOV-1999 (TREMBLrel. 12, Created)
DT 01-NOV-1999 (TREMBLrel. 12, Last sequence update)
DT 01-MAY-2000 (TREMBLrel. 13, Last annotation update)
DE ENVELOPE GLYCOPROTEIN (FRAGMENT).
GN ENV.
OS Human immunodeficiency virus type 1.
OC Viruses; Retroid viruses; Retroviridae; Lentivirus.
OX NCBI_TaxID=11676;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=THSC118;
RA Subbarao S., Vanichseni S., Hu D.;
"Genetic Characterization of Incident HIV-1 Subtype E and B Strains
from a Prospective Cohort of Injecting Drug Users in Bangkok,
Thailand.";
RL Submitted (MAY-1999) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF151752; AAD44294.1; -.
DR INTERPRO; IPR000777; -.
DR PFAM; PF00516; GP120; 1.
FT NON_TER 1
FT NON_TER 114
SQ SEQUENCE 114 AA; 13007 MW; 23C765F037B7CC5E CRC64;

Query Match 22.5%; Score 88.5; DB 14; Length 114;
Best Local Similarity 32.2%; Pred. No. 0.0044;
Matches 29; Conservative 12; Mismatches 30; Indels 19; Gaps 5;

QY 2 IEVVC-----NDRLGKVC---RVKCNDDTITGDKLKLIAAQTGRWNKIV-----LKKW 47
Db 17 VEINCTRPSNNTRTGVTIGPGVFTGTGIDIRKAYCEVNGTKWVKLVQVTEKLEH 76
QY 48 Y--TIFKDHVSLGDYEI---HDGMNLELY 72
Db 77 YNKTIIFQPPSGDLEITTHFNCRGEFFY 106

RESULT 7
O9JBC9 PRELIMINARY; PRT; 203 AA.
ID O9JBC9
AC O9JBC9;
DT 01-OCT-2000 (TREMBLrel. 15, Created)
DT 01-OCT-2000 (TREMBLrel. 15, Last sequence update)
DT 01-OCT-2000 (TREMBLrel. 15, Last annotation update)
DE ENVELOPE GLYCOPROTEIN (FRAGMENT).
GN ENV.
OS Human immunodeficiency virus type 1.
OC Viruses; Retroid viruses; Retroviridae; Lentivirus.
OX NCBI_TaxID=11676;
RN [1]
RP SEQUENCE FROM N.A.
RC MEDLINE-20027278; PubMed-10559367;
RA Shankarappa R., Margolick J.B., Gange S.J., Rodrigo A.G., Upchurch D.,
Farzadegan H., Gupta P., Rinaldo C.R., Learn G.H., He X., Huang X.-L.,
Mullins J.I.;
"Consistent viral evolutionary changes associated with the progression
of human immunodeficiency virus type 1 infection.";
RL J. Virol. 73:10489-10502(1999).
DR EMBL; AF138608; AAF32933.1; -.
FT NON_TER 1
FT NON_TER 203
SQ SEQUENCE 203 AA; 22963 MW; A2BE13A5B80C5CBA CRC64;

Query Match 22.3%; Score 88; DB 14; Length 203;
Best Local Similarity 37.1%; Pred. No. 0.01;
Matches 26; Conservative 7; Mismatches 23; Indels 14; Gaps 3;

QY 7 NDRLGKRVKVC-----NTDITGDKLKLIAAQTGRWNKIV-----YTIKF 52
Db 34 NRHTGKSIRNGLGRAWHTTTEIIGDIRKAYCTLNGTNWKNILEQVVKLVQVTEK 93
```

```

QY 53 DHVSLGDYEI 62
Db 94 NHSSGGDLI 103

RESULT 8
Q71412 PRELIMINARY; PRT; 115 AA.
AC Q71412;
DT 01-NOV-1996 (TReMBLrel. 01, Created)
DT 01-NOV-1996 (TReMBLrel. 01, Last sequence update)
DE 01-MAY-2000 (TReMBLrel. 13, Last annotation update)
DE ENVELOPE GLYCOPROTEIN, V3 REGION (FRAGMENT).
GN ENV.
OS Human immunodeficiency virus type 1.
OC Viruses; Retroid viruses; Retroviridae; Lentivirus.
OX NCBI_TaxID=11676;
RN [1]
RP SEQUENCE FROM N.A.
RX STRAIN=TH100;
RX MEDLINE=96014957; PubMed=7576318;
RA Kalish M.L., Baldwin A., Raktham S., Wasi C., Luo C., Schochetman G.,
RA Mastro T.D., Young N., Vanichseni S., Rubsamen-Waigmann H.,
RA Von Briesen H., Mullins J.I., Delwart E., Herring B., Esparza J.,
RA Heyward W.L., Osmanov S.;
RA "The evolving molecular epidemiology of HIV-1 envelope subtypes in
RA injecting drug users in Bangkok, Thailand: implications for HIV
RA vaccine trials.";
RL AIDS 9:851-857(1995).
RN [2]
RP SEQUENCE FROM N.A.
RX STRAIN=TH100;
RX MEDLINE=96014956; PubMed=7576317;
RA Wasi C., Herring B., Raktham S., Vanichseni S., Mastro T.D.,
RA Young N.L., Rubsamen-Waigmann H., Von Briesen H., Kalish M.L., Luo C.,
RA Pau C., Baldwin A., Mullins J.I., Delwart E.L., Esparza J.,
RA Heyward W.L., Osmanov S.;
RA "Determination of HIV-1 subtypes in injecting drug users in Bangkok,
RA Thailand, using peptide-binding enzyme immunoassay and heteroduplex
RA mobility assay: evidence of increasing infection with HIV-1 subtype
RA E.";
RL AIDS 9:843-849(1995).
RN [3]
RP SEQUENCE FROM N.A.
RX STRAIN=TH100;
RX SIEPEL A.C.;
RL Submitted (MAR-1995) to the EMBL/GenBank/DBJ databases.
DR EMBL; U22542; AAA88941.1; -.
DR INTERPRO; IPR000777; -.
DR PFAM; PF00516; GP120; 1.
ENVELOPE protein...
NON_TER 1 1
NON_TER 115 115
SQ SEQUENCE 115 AA; 13082 MW; 505DB9387E8A1918 CRC64;

Query Match 22.1%; Score 87; DB 14; Length 115;
Best Local Similarity 29.2%; Pred. No. 0.0067;
Matches 28; Conservative 7; Mismatches 27; Indels 34; Gaps 4;

QY 10 LGKRVKVC-----NTDDTIGDLKLLIAAQTGTRWKNKIVLKKWYT 49
Db 13 LNKSVENCTRPSNTRTSIRIGPGQVFKYTGDIIGDIRKAYCEINGTKWKNK-ALKQVAE 71
QY 50 IFKDHV-----SLGDYEI---HDGMNLELYY 72
Db 72 KEKEHFNNITTFQPPSGDLEITTHFNCRCGEFFY 107

RESULT 9
Q89553 PRELIMINARY; PRT; 216 AA.
ID 089553

```

```

AC 089553;
DT 01-NOV-1998 (TReMBLrel. 08, Created)
DT 01-NOV-1998 (TReMBLrel. 08, Last sequence update)
DE 01-MAY-2000 (TReMBLrel. 13, Last annotation update)
DE ENVELOPE GLYCOPROTEIN (FRAGMENT).
GN ENV.
OS Human immunodeficiency virus type 1.
OC Viruses; Retroid viruses; Retroviridae; Lentivirus.
OX NCBI_TaxID=11676;
RN [1]
RP SEQUENCE FROM N.A.
RX STRAIN=L6132;
RA Chen Y.-M., Lee C.-M.;
RA "Molecular epidemiology of HIV-1 infection in Taiwan.";
RL Submitted (JAN-1998) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF041121; AAC25046.1; -.
DR INTERPRO; IPR000777; -.
DR PFAM; PF00516; GP120; 1.
ENVELOPE protein.
NON_TER 1 1
NON_TER 216 216
SQ SEQUENCE 216 AA; 24532 MW; C5DDF30E6A65A6C7 CRC64;

Query Match 22.1%; Score 87; DB 14; Length 216;
Best Local Similarity 33.8%; Pred. No. 0.014;
Matches 26; Conservative 12; Mismatches 23; Indels 16; Gaps 4;

QY 2 LEVVC---NDRLGKKVR-----VKCNTDDTIGDLKLLIAAQTGTRWKNKIV-----LKKW 47
Db 13 VEINCTRPDSNVRTSIRIGPGQVFKYTGDIIGDIRKAYCEINGTKWKNKVLTVQTEKLEH 72
QY 48 Y--TIFKDRVSLGDYEI 62
Db 73 FNKTIFFQPPSGDLEI 89

RESULT 10
Q9WNN3 PRELIMINARY; PRT; 115 AA.
ID Q9WNN3
AC Q9WNN3;
DT 01-NOV-1999 (TReMBLrel. 12, Created)
DT 01-NOV-1999 (TReMBLrel. 12, Last sequence update)
DT 01-MAY-2000 (TReMBLrel. 13, Last annotation update)
DE ENVELOPE GLYCOPROTEIN (FRAGMENT).
GN ENV.
OS Human immunodeficiency virus type 1.
OC Viruses; Retroid viruses; Retroviridae; Lentivirus.
OX NCBI_TaxID=11676;
RN [1]
RP SEQUENCE FROM N.A.
RX STRAIN=THSC78;
RA Subbarao S., Vanichseni S., Hu D.;
RA "Genetic Characterization of Incident HIV-1 Subtype E and B Strains
RA from a Prospective Cohort of Injecting Drug Users in Bangkok,
RA Thailand.";
RL Submitted (MAY-1999) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF151775; AAD44317.1; -.
DR INTERPRO; IPR000777; -.
DR PFAM; PF00516; GP120; 1.
ENVELOPE protein.
NON_TER 1 1
NON_TER 115 115
SQ SEQUENCE 115 AA; 13144 MW; F6A683AFFC5A467B CRC64;

Query Match 21.8%; Score 86; DB 14; Length 115;
Best Local Similarity 28.9%; Pred. No. 0.0088;
Matches 28; Conservative 8; Mismatches 27; Indels 34; Gaps 4;

QY 9 RLGGKRVKVC-----NTDDTIGDLKLLIAAQTGTRWKNKIVLKKWY 48
Db 12 QLKNKSVENCTRPSNTRTSIRIGPGQVFKYTGDIIGDIRKAYCEINGTKWKNK-VLKQVT 70

```

QY 49 TIFKDHV-----SLGDYVEI---HDGMNLELY 72  
 Db 71 EKLKHEFNKTIIFQPPSGDLEITHHFNCKGFEFF 107  
 RESULT 11  
 O41876 PRELIMINARY; PRT; 164 AA.  
 ID O41876  
 AC O41876  
 DT 01-JAN-1998 (TREMBlrel. 05, Created)  
 DT 01-JAN-1998 (TREMBlrel. 05, Last sequence update)  
 DT 01-MAY-2000 (TREMBlrel. 13, Last annotation update)  
 DE ENVELOPE GLYCOPROTEIN (FRAGMENT).  
 GN ENV.  
 OS Human immunodeficiency virus type 1.  
 OC Viruses; Retroid viruses; Retroviridae; Lentivirus.  
 OX NCBI\_TaxID=11676;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=VN23;  
 RX MEDLINE=96344000; PubMed=8738437;  
 RA Nerurkar V.R., Nguyen H.T., Dashwood W.M., Hoffmann P.R., Yin C.,  
 RA Morens D.M., Kaplan A.H., Detels R., Yanagihara R.;  
 RT "HIV type 1 subtype E in commercial sex workers and injection drug  
 RT users in southern Vietnam.";  
 RL AIDS Res. Hum. Retroviruses 12:841-843(1996).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=VN23;  
 RA Nerurkar V.R., Nguyen H.T., Woodward C.L., Hoffmann P.R.,  
 RA Dashwood W.-M., Long H.T., Morens D.M., Detels R., Yanagihara R.;  
 RL Cell. Mol. Biol. 0:0-0(1997).  
 DR EMBL; U90087; AAB64376.1; -;  
 DR INTERPRO: IPR000777; -;  
 DR PFAM; PF00516; GP120; 1.  
 KW Envelope protein.  
 FT NON\_TER 1 1  
 FT NON\_TER 164 164  
 SQ SEQUENCE 164 AA; 18501 MW; 1D36498B47DA17C7 CRC64;

Query Match 21.8%; Score 86; DB 14; Length 164;  
 Best Local Similarity 28.4%; Pred. No. 0.013;  
 Matches 27; Conservative 10; Mismatches 26; Indels 32; Gaps 4;  
 10 LCKKVRVC-----NTDDTIGDLKLLIAAQTGRNKKIV-----43  
 Db 27 LKNSVEINCRPSNNRTSMTIGPGQVYRTGDIIGDKAYCEINGTKWNEVLKQVAGK 86

QY 44 LKKWY---TIFKDHVSLGDYVEI---HDGMNLELY 72  
 Db 87 LKHEFNKTIIFQPPSGDLEITHHFNCKGFEFF 121

RESULT 12  
 Q71495 PRELIMINARY; PRT; 114 AA.  
 ID Q71495  
 AC Q71495  
 DT 01-NOV-1996 (TREMBlrel. 01, Created)  
 DT 01-NOV-1996 (TREMBlrel. 01, Last sequence update)  
 DT 01-MAY-2000 (TREMBlrel. 13, Last annotation update)  
 DE ENVELOPE GLYCOPROTEIN, V3 REGION (FRAGMENT).  
 GN ENV.  
 OS Human immunodeficiency virus type 1.  
 OC Viruses; Retroid viruses; Retroviridae; Lentivirus.  
 OX NCBI\_TaxID=11676;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=TH98;  
 RX MEDLINE=96014957; PubMed=7576318;  
 RA Kalish M.L., Baldwin A., Raktham S., Wasi C., Luo C., Schochetman G.,  
 RA Mastro T.D., Young N., Vanichseni S., Rubsamen-Waigmann H.,

RA Von Briesen H., Mullins J.L., Delwart E., Herring B., Esparza J.,  
 RA Heyward W.L., Osmanov S.;  
 RT "The evolving molecular epidemiology of HIV-1 envelope subtypes in  
 RT injecting drug users in Bangkok, Thailand: implications for HIV  
 RT vaccine trials.";  
 RL AIDS 9:851-857(1995).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=TH98;  
 RX MEDLINE=96014956; PubMed=7576317;  
 RA Wasi C., Herring B., Raktham S., Vanichseni S., Mastro T.D.,  
 RA Young N.L., Rubsamen-Waigmann H., Von Briesen H., Kalish M.L., Luo C.,  
 RA Pau C., Baldwin A., Mullins J.L., Delwart E.L., Esparza J.,  
 RA Heyward W.L., Osmanov S.;  
 RT "Determination of HIV-1 subtypes in injecting drug users in Bangkok,  
 RT Thailand, using peptide-binding enzyme immunoassay and heteroduplex  
 RT mobility assay: evidence of increasing infection with HIV-1 subtype  
 RT E.";  
 RL AIDS 9:843-849(1995).  
 RN [3]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=TH98;  
 RA Siepel A.C.;  
 RL Submitted (MAR-1995) to the EMBL/GenBank/DBJ databases.  
 DR EMBL; U22625; AAA89024.1; -;  
 DR INTERPRO: IPR000777; -;  
 DR PFAM; PF00516; GP120; 1.  
 KW Envelope protein.  
 FT NON\_TER 1 1  
 FT NON\_TER 114 114  
 SQ SEQUENCE 114 AA; 13191 MW; 3801EA986E40539E CRC64;  
 Query Match 21.7%; Score 85.5; DB 14; Length 114;  
 Best Local Similarity 37.1%; Pred. No. 0.0099;  
 Matches 26; Conservative 6; Mismatches 25; Indels 13; Gaps 3;  
 QY 15 RVKCNITDDTIGDLKLLIAAQTGRNKKIVLKKWYIFKDHV-----SLGDYVEI---62  
 Db 38 RVYKGTGDIIGDKAYCEINGTKWKN-VLKQVTEKLEHFNRTILFQPPSGDLEITHM 96  
 QY 63 HDGMNLELY 72  
 Db 97 HFNCRGFEFF 106  
 RESULT 13  
 O90467 PRELIMINARY; PRT; 114 AA.  
 ID O90467  
 AC O90467  
 DT 01-NOV-1998 (TREMBlrel. 08, Created)  
 DT 01-NOV-1998 (TREMBlrel. 08, Last sequence update)  
 DT 01-MAY-2000 (TREMBlrel. 13, Last annotation update)  
 DE GP120 (FRAGMENT).  
 GN ENV.  
 OS Human immunodeficiency virus type 1.  
 OC Viruses; Retroid viruses; Retroviridae; Lentivirus.  
 OX NCBI\_TaxID=11676;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=THSC70;  
 RA Subbarao S., Vanichseni S., Kitavaporn D., Hart C., Raktham S.,  
 RA Young N., Srisuwanalai L., Luo C.C., Mastro T.D.;  
 RT "Genetic Characterization of Incident HIV-1 Subtype B and E Strains  
 RT from Injecting Drug Users (IDUs) in Bangkok, Thailand.";  
 RL Submitted (AUG-1998) to the EMBL/GenBank/DBJ databases.  
 DR EMBL; AF081761; AAC34019.1; -;  
 DR INTERPRO: IPR000777; -;  
 DR PFAM; PF00516; GP120; 1.  
 FT NON\_TER 1 1  
 FT NON\_TER 114 114  
 SQ SEQUENCE 114 AA; 12973 MW; F4CF55404ED58FC4 CRC64;



GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: June 13, 2001, 20:24:28 ; Search time 34.18 Seconds  
(without alignments)  
73.161 Million cell updates/sec

Title: US-09-331-930A-2

Perfect score: 394

Sequence: 1 MIEVVCNDRLGKVRKVCNT.....HVSIGDYEHGMMNLELYQ 73

Scoring table:

BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 93435 seqs, 34255486 residues

Total number of hits satisfying chosen parameters: 93435

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SwissProt\_39:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	73.5	18.7	949	HYSA_STRPN	Q54873 streptococ
2	65.5	16.6	504	CTB7_ARATH	Q98514 arabidopsis
3	65.5	16.6	797	CTPE_MYCTU	O08365 mycobacteri
4	65	16.5	333	Y642_METUA	Q58059 methanococ
5	62	15.7	527	RAG2_RABIT	P34089 oryctolagus
6	62	15.7	662	ABPA_STRLI	P53627 streptomyce
7	61	15.5	177	VC08_VACCV	P17364 vaccinia vi
8	60.5	15.4	338	ANX2_CHICK	P17785 gallus gall
9	60.5	15.4	1275	GNRP_HUMAN	Q13972 homo sapien
10	60	15.2	69	TRY1_SALTI	P12059 salmonella
11	60	15.2	740	PSAB_SYNEN	P25897 synchococc
12	59.5	15.1	644	YH0B_DROME	P12258 drosophila
13	59	15.0	669	ENV_MLVAV	P03386 akv murine
14	59	15.0	730	PSAB_SYNK3	P29255 synchocyst
15	59	15.0	732	PSAB_SYNP2	P17155 synchococc
16	58.5	14.8	330	CATS_RAT	Q02765 rattus norv
17	58.5	14.8	861	ENV_HV1KB	P31819 human immun
18	58	14.7	93	UBTL_NPVOP	O05120 orgyia pseu
19	58	14.7	184	VC08_VACCC	P21041 vaccinia vi
20	58	14.7	381	Y438_METUA	Q57880 methanococ
21	58	14.7	734	PSAB_PORPU	P51285 porphyra pu
22	57.5	14.6	185	PTH_RICPR	Q92cv4 rickettsia
23	57.5	14.6	339	ANX2_XENLA	P24801 xenopus lae
24	57.5	14.6	343	HRCA_MYCCE	P47447 mycoplasma
25	57.5	14.6	661	MXI_HUMAN	P20591 homo sapien
26	57.5	14.6	732	ERG7_HUMAN	P48449 homo sapien
27	57.5	14.6	1024	SYTP_STAAU	P41368 staphylococ
28	57	14.5	267	CB00_PSEHY	Q51858 pseudomonas
29	57	14.5	434	TB00_HUMAN	Q9867 homo sapien
30	56.5	14.3	222	ERG2_YEAST	P33352 saccharomyc
31	56.5	14.3	338	ANX2_MOUSE	P07356 mus musculu
32	56.5	14.3	468	SYFA_SULSO	P95961 sulfolobus
33	56	14.2	414	KIM8_HUMAN	Q76015 homo sapien

34 56 14.2 496 1 MGLA\_TREPA Q56342 treponema p  
35 56 14.2 734 1 PSAB\_GUITH O78507 guillardia  
36 56 14.2 1262 1 GNRP\_MOUSE P27671 mus musculu  
37 56 14.2 4486 1 DYH9\_HUMAN Q9nyc9 homo sapien  
38 55.5 14.1 261 1 RS3A\_ORYSA P49397 oryza sativ  
39 55.5 14.1 277 1 VA38\_VARV P33853 variola vir  
40 55.5 14.1 282 1 NADA\_APICA P29241 aplysia cal  
41 55.5 14.1 290 1 YEAB\_BACSU P46348 bacillus su  
42 55.5 14.1 867 1 SFMD\_ECOLI P77468 escherichia  
43 55.5 14.1 933 1 MSH2\_RAT P54275 rattus norv  
44 55.5 14.1 1008 1 VGLM\_UUK P09613 ukuniemi v  
45 55 14.0 267 1 LPXA\_PROMI P72215 proteus mir

#### ALIGNMENTS

RESULT 1

HYSA\_STRPN STANDARD; PRT; 949 AA.  
AC Q54873; Q54874;  
DT 15-DEC-1998 (Rel. 37, Created)  
DT 15-DEC-1998 (Rel. 37, Last sequence update)  
DT 15-DEC-1998 (Rel. 37, Last annotation update)  
DE HYALURONATE LYASE PRECURSOR (EC 4.2.2.1) (HYALURONIDASE) (HYASE).  
OS Streptococcus pneumoniae.  
OC Bacteria; Firmicutes; Bacillus/Clostridium group; Streptococcaceae;  
OC Streptococcus.  
OX NCBI\_TaxID=1313;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=TYPE 23;  
RX MEDLINE=94156460; PubMed=8112843;  
RA Berry A.M., Lock R.A., Thomas S.M., Rajan D.P., Hansman D.,  
Ratton J.C.;  
RT "Cloning and nucleotide sequence of the Streptococcus pneumoniae  
hyaluronidase gene and purification of the enzyme from recombinant  
Escherichia coli.";  
RL Infect. Immun. 62:1101-1108(1994).  
RN [2]  
RP CRYSTALLIZATION.  
RX MEDLINE=98234706; PubMed=9573623;  
RA Jedrejz M.J., Chantalat L., Mewbourne R.B.;  
RT "Crystallization and preliminary X-ray analysis of Streptococcus  
pneumoniae hyaluronate lyase.";  
RL J. Struct. Biol. 121:73-75(1998).  
CC -|- CATALYTIC ACTIVITY: HYALURONATE = N 3-(4-DEOXY-BETA-D-GLUC-4-  
ENURONOSYL)-N-ACETYL-D-GLUCOSAMINE.  
CC -|- SUBCELLULAR LOCATION: TYPE I MEMBRANE PROTEIN. CELL WALL.  
CC -|- SIMILARITY: BELONGS TO THE POLYSACCHARIDE LYASE FAMILY 8.  
CC -|- SIMILARITY: TO OTHER STREPTOCOCCAL AND STAPHYLOCOCCAL PROTEINS  
IN THE REGION OF THE MEMBRANE ANCHOR.  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
between the Swiss Institute of Bioinformatics and the EMBL outstation -  
the European Bioinformatics Institute. There are no restrictions on its  
use by non-profit institutions as long as its content is in no way  
modified and this statement is not removed. Usage by and for commercial  
entities requires a license agreement (See http://www.isb-sib.ch/announce/  
or send an email to license@isb-sib.ch).  
CC -----  
CC EMBL; L20670; AAA53685.1; -;  
DR EMBL; L20670; AAA53685.1; ALT\_INIT.  
DR InterPro; IPR001899; -;  
DR PROSITE; PS00343; GRAM\_POS\_ANCHORING; 1.  
DR Cell wall; Transmembrane; Lyase; Signal.  
FT SIGNAL 1 ? POTENTIAL.  
FT CHAIN ? 949 HYALURONATE LYASE.  
FT DOMAIN 1 924 EXTRACELLULAR (POTENTIAL).  
FT TRANSMEM 925 943 MEMBRANE ANCHOR (POTENTIAL).  
FT DOMAIN 944 949 CYTOPLASMIC (POTENTIAL).  
FT SEQUENCE 949 AA; 107756 MW; 77FE0AC7A765C3E2 CRC64;

```
Query Match      16.6%; Score 65.5; DB 1; Length 797;
Best Local Similarity 29.6%; Pred. No. 5.6;
Matches 24; Conservative 14; Mismatches 22; Indels 21; Gaps 5;
```



Streptomyces lividans.  
Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;  
Actinomycetales; Streptomycineae; Streptomycetaceae; Streptomyces.  
NCBI\_TaxID=1916;  
[1]  
SEQUENCE FROM N.A., AND SEQUENCE OF 1-23.  
STRAIN=66 / 1326;  
MEDLINE=94379974; PubMed=8092996;  
Manin C., Shareek F., Morosoli R., Kluepfel D.;  
"Purification and characterization of an alpha-L-arabinofuranosidase  
from Streptomyces lividans 66 and DNA sequence of the gene (abfa).";  
Biochem. J. 302:443-449(1994).  
-1- FUNCTION: IMPORTANT ROLE IN THE DEGRADATION OF ARABINOSE-  
CONTAINING HEMICELLULOSES. HYDROLYSES RAPIDLY THE SHORT-CHAIN  
ARABINO-OLIGOSACCHARIDES FROM DIGESTION OF XILAN WITH XILANASES.  
ACTS SLOWLY ON ARABINAN AND ARABINOXYLAN FROM WHEAT AND RYE FLOUR  
TO RELEASE L-ARABINOFURANOSE. DOES NOT ACT ON OAT-SPELT'S XILAN OR  
ARABINOGALACTAN. OPTIMAL ENZYME ACTIVITY OCCURS AT 60 DEGREES  
CELSIUS AND PH 6.0.  
-1- CATALYTIC ACTIVITY: HYDROLYSIS OF TERMINAL NON-REDUCING ALPHA-L-  
ARABINOFURANOSIDE RESIDUES IN ALPHA-L-ARABINOSIDES.  
-1- SUBUNIT: OLIGOMER.  
-1- SUBCELLULAR LOCATION: CYTOPLASMIC.  
-1- SIMILARITY: BELONGS TO FAMILY 51 OF GLYCOSYL HYDROLASES.  
-----  
This SWISS-PROT entry is copyright. It is produced through a collaboration  
between the Swiss Institute of Bioinformatics and the EMBL outstation -  
the European Bioinformatics Institute. There are no restrictions on its  
use by non-profit institutions as long as its content is in no way  
modified and this statement is not removed. Usage by and for commercial  
entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
-----  
EMBL; U04630; AAA61708.1; -  
Hydrolase; Glycosidase.  
SEQUENCE 662 AA; 72496 MW; DAAF66A577C1D6D1 CRC64;  
-----  
Query Match 15.7%; Score 62; DB 1; Length 662;  
Best Local Similarity 45.5%; Pred. No. 12;  
Matches 15; Conservative 4; Mismatches 12; Indels 2; Gaps 2;  
-----  
OY 34 QTGRNKKVILKWTIFKDHVSLDGYEIHG 65  
Db 221 ETPAEATVLTETIDLV-DHISLHAYEPHDG 252  
-----  
RESULT 7  
VC08\_VACCV STANDARD; PRT; 177 AA.  
ID VC08\_VACCV  
P17364;  
01-AUG-1990 (Rel. 15, Created)  
01-AUG-1990 (Rel. 15, Last sequence update)  
01-FEB-1991 (Rel. 17, Last annotation update)  
DE PROTEIN C8.  
GN C8.  
OS Vaccinia virus (strain WR).  
OC Viruses; dsDNA viruses, no RNA stage; Poxviridae; Chordopoxvirinae;  
OC Orthopoxvirus.  
OX NCBI\_TaxID=10254;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=89073756; PubMed=2849238;  
RA Kotwal G.J., Moss B.;  
RT "Analysis of a large cluster of nonessential genes deleted from a  
vaccinia virus terminal transposition mutant.";  
RL Virology 167:524-537(1988).  
CC  
-----  
This SWISS-PROT entry is copyright. It is produced through a collaboration  
between the Swiss Institute of Bioinformatics and the EMBL outstation -  
the European Bioinformatics Institute. There are no restrictions on its  
use by non-profit institutions as long as its content is in no way  
modified and this statement is not removed. Usage by and for commercial

entites requires a license agreement (See <http://www.isb-sib.ch/announce/>  
or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
-----  
EMBL; M22812; AAA69600.1; -  
PIR; I31829; WZVZA9.  
KW Early protein.  
SQ SEQUENCE 177 AA; 20753 MW; 8C536DD9F90F0065 CRC64;  
-----  
Query Match 15.5%; Score 61; DB 1; Length 177;  
Best Local Similarity 44.8%; Pred. No. 3.6;  
Matches 13; Conservative 4; Mismatches 10; Indels 2; Gaps 1;  
-----  
OY 43 VLKWTIFKDHVSLDGYEIHG 71  
Db 67 VMTKCAFPIDTVSV--YTSNDNLNIEFY 93  
-----  
RESULT 8  
ANX2\_CHICK STANDARD; PRT; 338 AA.  
ID ANX2\_CHICK  
P17785;  
DT 01-AUG-1990 (Rel. 15, Created)  
DT 01-AUG-1990 (Rel. 15, Last sequence update)  
DT 01-OCT-2000 (Rel. 40, Last annotation update)  
DE ANNEXIN II (LIPOCORTIN II) (CALPACTIN I HEAVY CHAIN) (CHROMOBINDIN 8)  
DE (P36) (PROTEIN I) (PLACENTAL ANTICOAGULANT PROTEIN IV) (PAP-IV).  
GN ANXA2 OR ANX2.  
OS Gallus gallus (Chicken).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Archosauria; Aves; Neognathae; Galliformes; Phasianinae;  
OC Gallus.  
OX NCBI\_TaxID=9031;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=90332421; PubMed=2143014;  
RA Gerke V., Koch W.;  
RT "The cDNA sequence of chicken annexin II.";  
RL Nucleic Acids Res. 18:4246-4246(1990).  
RN [2]  
RP SEQUENCE OF 1-69.  
RX MEDLINE=88296872; PubMed=2456953;  
RA Johnsson N., Johnsson K., Weber K.;  
RT "A discontinuous epitope on p36, the major substrate of src tyrosine-  
protein-kinase, brings the phosphorylation site into the  
neighbourhood of a consensus sequence for Ca2+/lipid-binding  
proteins.";  
RL FEBS Lett. 236:201-204(1988).  
RN [3]  
RP SEQUENCE OF 1-29.  
RX MEDLINE=89052656; PubMed=2973411;  
RA Johnsson N., Marriot G., Weber K.;  
RT "p36, the major cytoplasmic substrate of src tyrosine protein kinase,  
binds to its p11 regulatory subunit via a short amino-terminal  
amphipathic helix.";  
RL EMBL J. 7:2435-2442(1988).  
CC -1- FUNCTION: CALCIUM-REGULATED MEMBRANE-BINDING PROTEIN WHOSE  
AFFINITY FOR CALCIUM IS GREATLY ENHANCED BY ANIONIC PHOSPHOLIPIDS.  
IT BINDS TWO CALCIUM IONS WITH HIGH AFFINITY.  
CC -1- SUBUNIT: Tetramer of 2 light chains (P10 proteins) and 2 heavy  
chains (P36 proteins).  
CC -1- SUBCELLULAR LOCATION: IN THE LAMINA BENEATH THE PLASMA MEMBRANE.  
CC -1- DOMAIN: CONTAINS FOUR HOMOLOGOUS REPEATS WITH A CONSENSUS  
SEQUENCE COMMON TO ALL ANNEXIN PROTEINS. A PAIR OF THESE REPEATS  
MAY FORM ONE BINDING SITE FOR CALCIUM AND PHOSPHOLIPID.  
CC -1- MISCELLANEOUS: IT MAY CROSS-LINK PLASMA MEMBRANE PHOSPHOLIPIDS  
WITH ACTIN AND THE CYTOSKELETON AND BE INVOLVED WITH EXOCYTOSIS.  
CC -1- SIMILARITY: BELONGS TO THE ANNEXIN FAMILY.  
CC  
-----  
This SWISS-PROT entry is copyright. It is produced through a collaboration  
between the Swiss Institute of Bioinformatics and the EMBL outstation -  
the European Bioinformatics Institute. There are no restrictions on its  
use by non-profit institutions as long as its content is in no way



CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

CC EMBL; X53334; CAA37421.1; -  
CC PIR; S10501; LUCH2.  
CC HSP; P04083; LBO9.  
CC InterPro; IPR001464; -  
CC InterPro; IPR002389; -  
CC Pfam; PF00191; annexin; 4.  
CC PRINTS; PR00196; ANNEXIN.  
CC PRINTS; PR00198; ANNEXIN1.  
CC PROSITE; PS00223; ANNEXIN; 4.  
KW Annexin; Calcium/phospholipid-binding; Repeat; Phosphorylation.  
FT INIT\_MET 0 0  
FT DOMAIN 1 23 P10 BINDING SITE (POTENTIAL).  
FT REPEAT 41 101 ANNEXIN 1.  
FT REPEAT 113 173 ANNEXIN 2.  
FT REPEAT 198 258 ANNEXIN 3.  
FT REPEAT 273 333 ANNEXIN 4.  
FT MOD\_RES 23 23 PHOSPHORYLATION (BY SRC).  
FT MOD\_RES 25 25 PHOSPHORYLATION (BY PKC).  
SQ SEQUENCE 338 AA; 38509 MW; 4A869FE9B1983C30 CRC64;

Query Match 15.4%; Score 60.5; DB 1; Length 338;  
Best Local Similarity 29.4%; Pred. No. 8.3;  
Matches 25; Conservative 9; Mismatches 24; Indels 27; Gaps 4;

QY 1 MIEVVC---NDRGKVKRV-----KCNWDDTIGDLKLL-IAAQTGRNKKVILKK 46  
DB 127 LLEICSRNQLNEINRYREMYKTEKDIISDTSGDFRKLMLVALAKRC----- 179

QY 47 WYTFKDVHSLGDIYHIDGNLKY 71  
DB 180 -----EDTSVIDYELIQDARELY 198

RESULT 9  
ID GNRP\_HUMAN STANDARD; PRT; 1275 AA.  
AC Q13972;  
DT 01-OCT-2000 (Rel. 40, Created)  
DT 01-OCT-2000 (Rel. 40, Last sequence update)  
DT 01-OCT-2000 (Rel. 40, Last annotation update)  
DE GUANINE NUCLEOTIDE RELEASING PROTEIN (GNRP) (RAS-SPECIFIC NUCLEOTIDE  
EXCHANGE FACTOR CDC25).  
OS Homo sapiens (Human).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
OX NCBI\_TaxID=9606;  
RN [1]  
RP SEQUENCE FROM N.A.  
RA Wei W., Broek D.;  
RT "Cloning and analysis of the full length human cdc25 cDNA, a ras-  
specific nucleotide exchange factor."  
RL Submitted (DEC-1993) to the EMBL/GenBank/DBJ databases.  
CC -!- FUNCTION: PROMOTES THE EXCHANGE OF RAS-BOUND GDP BY GTP.  
CC -!- SIMILARITY: CONTAINS 2 PH DOMAINS.  
CC -!- SIMILARITY: CONTAINS 1 DBL-HOMOLOGY DOMAIN (DH).  
CC -!- SIMILARITY: CONTAINS 1 RASCEF DOMAIN.  
CC -!- SIMILARITY: CONTAINS 1 IQ DOMAIN.  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

CC EMBL; L26584; AAA58417.1; -

DR InterPro; IPR000048; -  
DR InterPro; IPR000219; -  
DR InterPro; IPR000651; -  
DR InterPro; IPR001331; -  
DR InterPro; IPR001849; -  
DR InterPro; IPR001895; -  
DR Pfam; PF00612; IQ; 1.  
DR Pfam; PF00169; PH; 2.  
DR Pfam; PF00617; RasGEF; 1.  
DR Pfam; PF00618; RasGEFN; 1.  
DR Pfam; PF00621; RhoGEF; 1.  
DR PROSITE; PS00720; GDS\_CDC25; 1.  
DR PROSITE; PS00741; GDS\_CDC24; 1.  
DR PROSITE; PS00003; PH\_DOMAIN; 2.  
DR PROSITE; PS00096; IQ; 1.  
KW Guanine-nucleotide releasing factor.  
FT DOMAIN 22 129 PH 1.  
FT DOMAIN 204 229 IQ.  
FT DOMAIN 247 431 DH.  
FT DOMAIN 467 584 PH 2.  
FT DOMAIN 1038 1272 RASGEF.  
SQ SEQUENCE 1275 AA; 145381 MW; 86C6F54AA1E451F1 CRC64;

Query Match 15.4%; Score 60.5; DB 1; Length 1275;  
Best Local Similarity 32.0%; Pred. No. 36;  
Matches 16; Conservative 12; Mismatches 17; Indels 5; Gaps 3;

QY 7 NDRGKVKRVKCNWDDTIGDLKLLIAAQTGR-WNKVILKKWYTFKDHV 55  
DB 185 NERIQTQTVAPEDEDS--DIKKIKKQVSLRGW--LCRRKWKTIQDYI 230

RESULT 10  
TRY1\_SALTY STANDARD; PRT; 69 AA.  
ID TRY1\_SALTY STANDARD; PRT; 69 AA.  
AC P12059;  
DT 01-OCT-1989 (Rel. 12, Created)  
DT 01-OCT-1989 (Rel. 12, Last sequence update)  
DT 15-DEC-1998 (Rel. 37, Last annotation update)  
DE TRAY PROTEIN.  
GN TRAY.  
OS Salmonella typhi.  
OC Plasmid IncFv pBD208.  
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;  
OC Salmonella.  
OX NCBI\_TaxID=601;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=87056998; PubMed=2877970;  
RA Finlay B.B., Frost L.S., Paranchych W.;  
RT "Nucleotide sequence of the tra YALE region from IncFv plasmid  
pED208."  
RL J. Bacteriol. 168:990-998(1986).  
CC -!- FUNCTION: INVOLVED IN THE CONJUGATION PROCESS OF BACTERIAL CELLS  
CC FOR THE EXCHANGE OF PLASMID DNA. IT IS ALSO RESPONSIBLE FOR  
CC CONJUGAL DNA METABOLISM. TRAY IS REQUIRED FOR STRAND-SPECIFIC  
CC NICKING AT ORIT, THE TRANSFER ORIGIN.  
CC -!- SIMILARITY: 26%, 32% AND 25% SIMILARITY TO TRAY PROTEINS OF  
CC PLASMID F, R1 AND R100, RESPECTIVELY IN THE REGION OF POS. 10-60.  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

CC EMBL; M14733; AAA25604.1; -  
CC EMBL; M14733; AAA25605.1; ALT\_TERM.  
CC EMBL; M14733; AAA25606.1; ALT\_INIT.  
CC PIR; A25161; A25161.



Db 340 KKIYKCKSNLQHTLETTLSACPTRWNSNYKMKMSILDNRWSDVK---ILGEADIHVD 396

QY 66 MN 67

Db 397 FN 398

RESULT 13

ENV\_MLVAV STANDARD; PRT; 669 AA.

ID ENV\_MLVAV AC P03386;

DT 21-JUL-1986 (Rel. 01, Created)

DT 21-JUL-1986 (Rel. 01, Last sequence update)

DT 01-OCT-1996 (Rel. 34, Last annotation update)

DE ENV POLYPROTEIN PRECURSOR (COAT POLYPROTEIN) [CONTAINS: KNOB PROTEIN

DE GP70; SPIKE PROTEIN P15E; R PROTEIN].

GN ENV.

OS AKV murine leukemia virus.

OS Viruses; Retroviral viruses; Retroviridae; Mammalian type C retroviruses.

NCBI\_TaxID=11791;

[1]

SEQUENCE FROM N.A.

RP MEDLINE=84115072; PubMed=6319746;

RA Herr W.;

RT "Nucleotide sequence of AKV murine leukemia virus.";

RL J. Virol. 49:471-478(1984).

[2]

SEQUENCE FROM N.A.

RP MEDLINE=83090450; PubMed=6294621;

RA Herr W.;

RT "Nucleotide sequence of the 3' half of AKV.";

RL Nucleic Acids Res. 10:6931-6944(1982).

[3]

SEQUENCE FROM N.A.

RP MEDLINE=82217034; PubMed=62831170;

RA Lenz J., Crowther R., Straceski A., Haseltine W.;

RT "Nucleotide sequence of the AKV env gene.";

RL J. Virol. 42:519-529(1982).

CC -1- PTM: SPECIFIC ENZYMIC CLEAVAGES IN VIVO YIELD MATURE PROTEINS.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration

CC between the Swiss Institute of Bioinformatics and the EMBL outstation

CC the European Bioinformatics Institute. There are no restrictions on its

CC use by non-profit institutions as long as its content is in no way

CC modified and this statement is not removed. Usage by and for commercial

CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>

CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

DR EMBL; J01998; AAB03092.1; -

DR EMBL; V01164; CAA24493.1; ALT\_INIT.

DR PIR; A03984; VCVWEK.

DR HSP; P03385; LMOF.

DR InterPro; IPR002050; -

DR Pfam; PF00429; ENV\_polyprotein; 1.

KW Coat protein; Glycoprotein; Polyprotein; Transmembrane; Signal.

FT SIGNAL 1 31

FT CHAIN 32 470 KNOB PROTEIN GP70.

FT CHAIN 471 650 SPIKE PROTEIN P15E.

FT CHAIN 651 669 R PROTEIN.

FT TRANSMEM 473 493 POTENTIAL.

FT TRANSMEM 612 632 POTENTIAL.

FT CARBOHYD 43 43 N-LINKED (GLCNAC. .) (POTENTIAL).

FT CARBOHYD 199 199 N-LINKED (GLCNAC. .) (POTENTIAL).

FT CARBOHYD 293 293 N-LINKED (GLCNAC. .) (POTENTIAL).

FT CARBOHYD 327 327 N-LINKED (GLCNAC. .) (POTENTIAL).

FT CARBOHYD 359 359 N-LINKED (GLCNAC. .) (POTENTIAL).

FT CARBOHYD 366 366 N-LINKED (GLCNAC. .) (POTENTIAL).

FT CARBOHYD 399 399 N-LINKED (GLCNAC. .) (POTENTIAL).

FT CONFLICT 35 35 G -> R (IN REF. 2).

FT CONFLICT 463 463 E -> K (IN REF. 2).

FT CONFLICT 592 592 E -> K (IN REF. 2).

SEQUENCE 669 AA; 73756 MW; 3AGC3845208A13F2 CRC64;

Query Match 15.0%; Score 59; DB 1; Length 669;

Best Local Similarity 37.1%; Pred. No. 26;

Matches 13; Conservative 4; Mismatches 18; Indels 0; Gaps 0;

QY 5 VCNDRGLGKVRVKCNTDDTTCGLDKKLIAAQTGTW 39

Db 385 ICIGAVPRTHQVLCNTTKTSQSGSYLAAPTGTW 419

RESULT 14

PSAB\_SYNV3 STANDARD; PRT; 730 AA.

ID PSAB\_SYNV3 AC P29255; P73398;

DT 01-DEC-1992 (Rel. 24, Created)

DT 01-NOV-1997 (Rel. 35, Last sequence update)

DT 01-NOV-1997 (Rel. 35, Last annotation update)

DE PHOTOSYSTEM I P700 CHLOROPHYLL A APOPROTEIN A2.

GN PSAB OR SLR1835.

OS Synechocystis sp. (strain PCC 6803).

OC Bacteria; Cyanobacteria; Chroococcales; Synechocystis.

OX NCBI\_TaxID=1148;

RN [1]

SEQUENCE FROM N.A.

RP MEDLINE=92032774; PubMed=1932686;

RA Smart L.B., McIntosh L.;

RT "Expression of photosynthesis genes in the cyanobacterium

RT Synechocystis sp. PCC 6803: psaa-psab and psba transcripts accumulate

RT in dark-grown cells.";

RL Plant Mol. Biol. 17:959-971(1991).

RN [2]

SEQUENCE FROM N.A.

RP MEDLINE=97061201; PubMed=8905231;

RA Kaneko T., Sato S., Kotani H., Tanaka A., Asamizu E., Nakamura Y.,

RA Miyajima N., Hirose M., Sugita M., Sasamoto S., Kimura T.,

RA Hosouchi T., Shimono A., Muraki A., Nakazaki N., Naruo K.,

RA Okumura S., Shimpo S., Takeuchi C., Wada T., Watanabe A.,

RA Yamada M., Yasuda M., Tabata S.;

RT "Sequence analysis of the genome of the unicellular cyanobacterium

RT Synechocystis sp. strain PCC6803. II. Sequence determination of the

RT entire genome and assignment of potential protein-coding regions.";

RL DNA Res. 3:109-136(1996).

CC -1- FUNCTION: PSAA ALONG WITH PSAB BIND P700, THE PRIMARY ELECTRON

CC DONOR OF PHOTOSYSTEM I (PSI) AS WELL AS THE ELECTRON ACCEPTORS A0,

CC A1, AND FX.

CC -1- COFACTOR: P700 IS A CHLOROPHYL DIMER, A0 IS CHLOROPHYLL, A1 IS A

CC PHYLLOQUINONE AND FX IS A 4FE-4S IRON-SULFUR CENTER.

CC -1- SUBUNIT: HETERODIMER OF A1 AND A2 SUBUNITS WHICH BIND TOGETHER THE

CC IRON-SULFUR CENTER FX.

CC -1- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN.

CC -1- SIMILARITY: BELONGS TO THE PSAA/PSAB FAMILY.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration

CC between the Swiss Institute of Bioinformatics and the EMBL outstation

CC the European Bioinformatics Institute. There are no restrictions on its

CC use by non-profit institutions as long as its content is in no way

CC modified and this statement is not removed. Usage by and for commercial

CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>

CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

DR EMBL; X58825; CAA41630.1; -

DR EMBL; D90906; BAA17438.1; -

DR PIR; S18243; S18243.

DR PIR; S19090; S19090.

DR InterPro; IPR001280; -

DR Pfam; PF00223; psaa-psab; 1.

DR PRINTS; PR00257; PHOTOSYSTEM\_I\_PSAAB; 1.

DR PROSITE; PS00419; PHOTOSYSTEM\_I\_PSAAB; 1.

KW Chlorophyll; Photosystem I; Electron transport; Photosynthesis;

KW Transmembrane; Iron-sulfur; 4Fe-4S.

FT INIT\_MET 0 0 BY SIMILARITY.

FT TRANSMEM 45 68 I (POTENTIAL).

FT TRANSMEM 134 156 II (POTENTIAL).

```

FT TRANSMEM 174 198 III (POTENTIAL).
FT TRANSMEM 272 290 IV (POTENTIAL).
FT TRANSMEM 327 350 V (POTENTIAL).
FT TRANSMEM 366 392 VI (POTENTIAL).
FT TRANSMEM 414 436 VII (POTENTIAL).
FT TRANSMEM 513 531 VIII (POTENTIAL).
FT TRANSMEM 571 592 IX (POTENTIAL).
FT TRANSMEM 639 661 X (POTENTIAL).
FT TRANSMEM 703 723 XI (POTENTIAL).
FT METAL 555 IRON-SULFUR (4FE-4S).
FT METAL 564 IRON-SULFUR (4FE-4S).
FT CONFLICT 536 537 DA -> ES (IN REF. 1).
SQ SEQUENCE 730 AA; 81160 MW; EC804AA9F80E1231 CRC64;

Query Match 15.0%; Score 59; DB 1; Length 730;
Best Local Similarity 48.1%; Pred. No. 29;
Matches 13; Conservative 1; Mismatches 9; Indels 4; Gaps 1;

OY 47 WYTIKDHVSLGDIYHDMNLELYQ 73
II I I I I I I I I I I
21 WYGIATAH----DFETHDGMTEENLYQ 43

RESULT 15
PSAB_SYN2 STANDARD; PRT; 732 AA.
AC P17155;
DT 01-AUG-1990 (Rel. 15, Created)
DT 01-JUL-1993 (Rel. 26, Last sequence update)
DT 01-JUN-1994 (Rel. 29, Last annotation update)
DE PHOTOSYSTEM I P700 CHLOROPHYLL A APOPROTEIN A2.
GN PSAB.
OS Synechococcus sp. (strain PCC 7002) (Agmenellum quadruplicatum).
OC Bacteria; Cyanobacteria; Chroococcales; Synechococcus.
OX NCBI_TaxID=32049;
RN [1]
RP SEQUENCE FROM N.A.
RA Cantrell A., Bryant D.A.;
RT "Molecular cloning and nucleotide sequence of the psaa and psab genes
of the cyanobacterium Synechococcus sp. PCC 7002.";
RL Plant Mol. Biol. 9:453-468(1987).
CC -I- FUNCTION: PSAA ALONG WITH PSAB BIND P700, THE PRIMARY ELECTRON
CC DONOR OF PHOTOSYSTEM I (PSI) AS WELL AS THE ELECTRON ACCEPTORS A0,
CC A1, AND FX.
CC -I- COFACTOR: P700 IS A CHLOROPHYL DIMER, A0 IS CHLOROPHYLL, A1 IS A
CC PHYLLOQUINONE AND FX IS A 4FE-4S IRON-SULFUR CENTER.
CC -I- SUBUNIT: HETERODIMER OF A1 AND A2 SUBUNITS WHICH BIND TOGETHER THE
CC IRON-SULFUR CENTER FX.
CC -I- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN.
CC -I- SIMILARITY: BELONGS TO THE PSAA/PSAB FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; M18165; AAA88634.1; -
DR PIR; S06902; S06902.
DR InterPro; IPR001280; -
DR Pfam; PF00223; psaa.psaab; 1.
DR PRINTS; PR00257; PHOTOSYPSAAB
DR PROSITE; PS00419; PHOTOSYSTEM_I_PSAAB; 1.
KW Chlorophyll; Photosystem I; Electron transport; Photosynthesis;
KW Transmembrane; Iron-sulfur; 4Fe-4S.
FT INIT_MET 0 0 BY SIMILARITY.
FT TRANSMEM 45 68 I (POTENTIAL).
FT TRANSMEM 134 156 II (POTENTIAL).
FT TRANSMEM 174 198 III (POTENTIAL).
FT TRANSMEM 272 290 IV (POTENTIAL).

```

```

FT TRANSMEM 329 352 V (POTENTIAL).
FT TRANSMEM 368 394 VI (POTENTIAL).
FT TRANSMEM 416 438 VII (POTENTIAL).
FT TRANSMEM 515 533 VIII (POTENTIAL).
FT TRANSMEM 573 594 IX (POTENTIAL).
FT TRANSMEM 641 663 X (POTENTIAL).
FT TRANSMEM 705 725 XI (POTENTIAL).
FT METAL 557 IRON-SULFUR (4FE-4S).
FT METAL 566 IRON-SULFUR (4FE-4S).
SQ SEQUENCE 732 AA; 81557 MW; 94473FBFDC200EA1 CRC64;

Query Match 15.0%; Score 59; DB 1; Length 732;
Best Local Similarity 48.1%; Pred. No. 29;
Matches 13; Conservative 1; Mismatches 9; Indels 4; Gaps 1;

OY 47 WYTIKDHVSLGDIYHDMNLELYQ 73
II I I I I I I I I I I
21 WYGIATAH----DFETHDGMTEENLYQ 43

Search completed: June 13, 2001, 20:27:53
Job time: 205 sec

```

GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: June 13, 2001, 19:30:07 ; Search time 178.32 Seconds  
(without alignments)  
1119.634 Million cell updates/sec

Title: US-09-331-930A-1  
Perfect score: 342  
Sequence: 1 gttccaggagattacagctc.....attttttgttaagtgtact 342

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Sorted: 678276 seqs, 291890651 residues

Total number of hits satisfying chosen parameters: 1356552

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : N\_Geneseq\_0401.\*  
1: /SID56/gcgdata/geneseq/geneseq/NA1980.DAT.\*  
2: /SID56/gcgdata/geneseq/geneseq/NA1981.DAT.\*  
3: /SID56/gcgdata/geneseq/geneseq/NA1982.DAT.\*  
4: /SID56/gcgdata/geneseq/geneseq/NA1983.DAT.\*  
5: /SID56/gcgdata/geneseq/geneseq/NA1984.DAT.\*  
6: /SID56/gcgdata/geneseq/geneseq/NA1985.DAT.\*  
7: /SID56/gcgdata/geneseq/geneseq/NA1986.DAT.\*  
8: /SID56/gcgdata/geneseq/geneseq/NA1987.DAT.\*  
9: /SID56/gcgdata/geneseq/geneseq/NA1988.DAT.\*  
10: /SID56/gcgdata/geneseq/geneseq/NA1989.DAT.\*  
11: /SID56/gcgdata/geneseq/geneseq/NA1990.DAT.\*  
12: /SID56/gcgdata/geneseq/geneseq/NA1991.DAT.\*  
13: /SID56/gcgdata/geneseq/geneseq/NA1992.DAT.\*  
14: /SID56/gcgdata/geneseq/geneseq/NA1993.DAT.\*  
15: /SID56/gcgdata/geneseq/geneseq/NA1994.DAT.\*  
16: /SID56/gcgdata/geneseq/geneseq/NA1995.DAT.\*  
17: /SID56/gcgdata/geneseq/geneseq/NA1996.DAT.\*  
18: /SID56/gcgdata/geneseq/geneseq/NA1997.DAT.\*  
19: /SID56/gcgdata/geneseq/geneseq/NA1998.DAT.\*  
20: /SID56/gcgdata/geneseq/geneseq/NA1999.DAT.\*  
21: /SID56/gcgdata/geneseq/geneseq/NA2000.DAT.\*  
22: /SID56/gcgdata/geneseq/geneseq/NA2001.DAT.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	342	100.0	342	20	X57359 P. obesus beacon D
2	342	100.0	342	21	C81767 Israeli sand rat b
3	235.2	68.8	419	20	X85745 Novel cDNA sequenc
4	205.6	60.1	413	21	C08654 Human secreted pro
5	205.6	60.1	427	21	C08655 Human secreted pro
6	197	57.6	502	21	C08656 Human secreted pro
7	184.2	53.9	333	16	T19132 Human gene signatu
8	180.4	52.7	315	16	T25014 Human gene signatu
9	123	36.0	624	21	C46615 zea mays DNA fragm
10	118	34.5	556	21	C47007 Arabidopsis thalia
11	111.2	32.5	472	21	C41484 zea mays DNA fragm

12	104.6	30.6	460	21	A31562 Plant microsatelli
13	77	22.5	102	20	X57370 Human beacon DNA.
14	77	22.5	102	21	C81770 Human beacon (shor
15	71	20.8	417	21	C08657 Human secreted pro
16	65.8	19.2	2180	21	C81769 Human beacon codin
17	64.2	18.8	311	21	A31226 Plant microsatelli
18	35.4	10.4	12492	8	N70543 Complete sequence
19	34.8	10.2	484	21	C41348 zea mays DNA fragm
20	33.8	9.9	442	21	C36774 Arabidopsis thalia
21	33.8	9.9	501	21	C37049 Arabidopsis thalia
22	33.8	9.9	731	21	C43835 Arabidopsis thalia
23	33.8	9.9	976	21	C51382 Arabidopsis thalia
24	33.8	9.9	1106	21	C39920 Arabidopsis thalia
25	33.8	9.9	1462	21	C48056 Arabidopsis thalia
26	33.4	9.8	3083	21	C62751 Super ubiquitin pr
27	32.6	9.5	424	20	X84148 DNA encoding human
28	32.6	9.5	424	21	C79377 3' cDNA sequence o
29	32.4	9.5	903	21	C34429 Arabidopsis thalia
30	32.2	9.4	520	21	C41013 zea mays DNA fragm
31	32.2	9.4	773	21	C41828 Arabidopsis thalia
32	32.2	9.4	3840	10	N92365 Sequence of maize
33	32	9.4	601	21	F11308 Aspergillus niger
34	31.8	9.3	1401	21	C44286 Arabidopsis thalia
35	31.8	9.3	1757	14	Q49428 Cytochrome p450 ho
36	31.8	9.3	1757	18	T94666 Petunia flavonoid
37	31.6	9.2	1087	21	C67872 Neospora caninum N
38	31.6	9.2	1087	21	C67872 Neospora caninum N
39	31.4	9.2	1965	21	C61108 Degenerate DNA seq
40	31	9.1	49999	20	Z23901 Human LOBO homolog
41	30.8	9.0	489	20	V86109 EST clone GI. Hom
42	30.6	8.9	1385	21	C51389 Arabidopsis thalia
43	30.4	8.9	1124	21	C44836 Arabidopsis thalia
44	30.4	8.9	2343	21	C61107 Human DNA encoding
45	30.4	8.9	5512	20	Z28435 Sugar cane ubiquit

#### ALIGNMENTS

RESULT 1  
X57359  
ID X57359 standard; DNA; 342 BP.

XX X57359;

XX 24-JUL-1999 (first entry)

XX P. obesus beacon DNA.

XX Beacon; hypothalamus; obese; lean; agonist; antagonist; treatment;  
XX obesity; anorexia; weight maintenance; energy imbalance; diabetes;  
XX metabolic syndrome; dyslipidemia; hypertension; insulin resistance;  
XX medication; livestock; diagnosis; ss.

XX Psammomys obesus.

XX Key Location/Qualifiers

XX CDS 29..250

XX FT /\*tag= a

XX FT /product= "beacon"

XX WO9923217-A1.

XX 14-MAY-1999.

XX 30-OCT-1998; 98WO-AU00902.

XX 11-NOV-1997; 97AU-0000323.

XX 31-OCT-1997; 97AU-0000117.

XX (ITDI-) INT DIABETES INST.

XX (UYDE-) UNIV DEAKIN.

PI Collier G, Zimmet PZ;  
XX WPI: 1999-337484/28.  
DR P-PSDB; Y08413.  
XX  
PT New gene encoding a beacon protein associated with modulation of  
PT obesity, diabetes and metabolic energy levels  
XX  
PS Claim 3; Page 49; 85pp; English.  
XX  
CC This invention describes a novel beacon protein and its encoding nucleic  
CC acid which is expressed in larger amounts in hypothalamus tissue of obese  
CC animals compared to lean animals. Agonists and antagonists of beacon can  
CC be used to treat obesity, anorexia, weight maintenance, energy imbalance,  
CC diabetes, metabolic syndrome, dyslipidemia, hypertension and/or insulin  
CC resistance. The beacon protein, itself is used to manufacture medicaments  
CC for treatment of obesity, anorexia, energy imbalance or diabetes. The  
CC treatment is contemplated for both human and animals, such as those  
CC important to the livestock industry. The antibody and polynucleotides are  
CC useful in diagnosis of conditions as above.  
XX  
Sequence 342 BP; 93 A; 76 C; 79 G; 94 T; 0 other;  
Query Match 100.0%; Score 342; DB 20; Length 342;  
Best Local Similarity 100.0%; Pred. No. 2.9e-106;  
Matches 342; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 1 gtccaggagattacagctccagccacacatgattgagtggtttgcaacgacgctctagg 60  
Db |||||||  
QY 61 aaagaagtcgcgcttaagtgcacacccagatgacaccatcggtggactggaagaaactgat 120  
Db |||||||  
QY 121 agcgcccaactgcactcgttgaaataagatcgtttcttaaaagtgtgacacatttt 180  
Db |||||||  
QY 181 taaggaccatgtatctctgggagattatgaaatccacgattggatgaaacctggagcttta 240  
Db |||||||  
QY 241 ttaccagtagaggggaattctccacctgcccacacctgttcttccctcccatggctca 300  
Db |||||||  
QY 301 tttaacactgttagatgctcattttttttttaaagtgtact 342  
Db |||||||  
RESULT 2  
C81767  
ID C81767 standard; cDNA; 342 BP.  
XX C81767;  
XX  
XX 23-FEB-2001 (first entry)  
DE Israel sand rat beacon coding sequence.  
XX  
KW Israel sand rat; beacon; obesity; NIDDM; energy balance; diabetes;  
KW ligand; ss.  
XX  
OS Psammomys obesus.  
XX  
PN WO200064931-A1.  
XX  
PD 02-NOV-2000.  
XX  
PF 19-APR-2000; 2000WO-AU00342.

XX 23-APR-1999; 99AU-0009919.  
PR WPI: 1999-337484/28.  
XX P-PSDB; Y08413.  
XX  
PA (AUTO-) AUTOGEN PTY LTD.  
XX  
XX Collier G, Walder K, Zimmet P;  
XX  
XX WPI: 2000-687311/67.  
DR P-PSDB; B36290.  
XX  
XX Ligand of beacon protein useful for treating obesity, anorexia, energy  
XX imbalance, diabetes, metabolic syndrome, dyslipidemia, hypertension and  
XX insulin resistance -  
XX  
XX Claim 2; Fig 1; 67pp; English.  
XX  
XX The present invention is related to the isolation of a ligand known as  
XX beacon from the Israeli sand rat. Beacon is associated with the  
XX regulation of energy balance, and the protein, its coding sequence and  
XX analogues can be used in the treatment of diabetes, obesity, anorexia,  
XX energy imbalance, metabolic syndrome, dyslipidemia, hypertension and  
XX insulin resistance. In addition, they can be used in agriculture to  
XX produce leaner animals.  
XX  
XX Sequence 342 BP; 93 A; 76 C; 79 G; 94 T; 0 other;  
SQ  
Query Match 100.0%; Score 342; DB 21; Length 342;  
Best Local Similarity 100.0%; Pred. No. 2.9e-106;  
Matches 342; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 1 gtccaggagattacagctccagccacacatgattgagtggtttgcaacgacgctctagg 60  
Db |||||||  
QY 61 aaagaagtcgcgcttaagtgcacacccagatgacaccatcggtggactggaagaaactgat 120  
Db |||||||  
QY 121 agcgcccaactgcactcgttgaaataagatcgtttcttaaaagtgtgacacatttt 180  
Db |||||||  
QY 181 taaggaccatgtatctctgggagattatgaaatccacgattggatgaaacctggagcttta 240  
Db |||||||  
QY 241 ttaccagtagaggggaattctccacctgcccacacctgttcttccctcccatggctca 300  
Db |||||||  
QY 301 tttaacactgttagatgctcattttttttttaaagtgtact 342  
Db |||||||  
RESULT 3  
X85745  
ID X85745 standard; cDNA; 419 BP.  
XX  
XX X85745;  
XX  
XX 06-SEP-1999 (first entry)  
DE Novel cDNA sequence from a mouse blastocyst cDNA library.  
XX  
XX Mouse; blastocyst; cDNA library; ss.  
XX  
XX Mus sp.  
XX  
XX JP11164691-A.  
XX









Db 61 tgataccatcggtgacaccttaagaagctgattgcagcccaaacctgtaccctgttgaaacaa 120  
Qy 151 gatcgtttcttaaaaaagtggtacacagatttttaaggaccattatctctggagattatga 210  
D 121 gatgtctctgaagaagtggtacacagatttttggggaccacgtgtctctggggactatga 180  
Qy 211 aatcacagtggtgatgaacctggagctttattaccagtagagggaattc 260  
D 181 aatcacagtggtgatgaacctggagctttattaccagtagagggaattc 230

RESULT 9

C46615

ID C46615 standard; DNA; 624 BP.

XX C46615;

AC C46615;

XX C46615;

DT 18-OCT-2000 (first entry)

DE Zea mays DNA fragment SEQ ID NO: 50789.

XX

Hybridisation assay; genetic mapping; gene expression control;  
protein identification; signal transduction pathway; metabolic;  
pathway; promoter; termination sequence; corn; ss.

XX

OS Zea mays subsp. mays.

XX

FN EP1033405-A2.

XX

XX

PD 06-SEP-2000.

XX

PF 25-FEB-2000; 2000EP-0301439.

XX

XX 25-FEB-1999; 99US-0121825.

PR

PR 05-MAR-1999; 99US-0123180.

PR

PR 09-MAR-1999; 99US-0123548.

PR

PR 23-MAR-1999; 99US-0125788.

PR

PR 25-MAR-1999; 99US-0126284.

PR

PR 29-MAR-1999; 99US-0126785.

PR

PR 01-APR-1999; 99US-0127462.

PR

PR 06-APR-1999; 99US-0128234.

PR

PR 08-APR-1999; 99US-0128714.

PR

PR 16-APR-1999; 99US-0129845.

PR

PR 19-APR-1999; 99US-0130077.

PR

PR 21-APR-1999; 99US-0130449.

PR

PR 23-APR-1999; 99US-0130510.

PR

PR 23-APR-1999; 99US-0130891.

PR

PR 28-APR-1999; 99US-0131449.

PR

PR 30-APR-1999; 99US-0132048.

PR

PR 30-APR-1999; 99US-0132407.

PR

PR 04-MAY-1999; 99US-0132484.

PR

PR 05-MAY-1999; 99US-0132485.

PR

PR 06-MAY-1999; 99US-0132486.

PR

PR 06-MAY-1999; 99US-0132487.

PR

PR 07-MAY-1999; 99US-0132863.

PR

PR 11-MAY-1999; 99US-0134236.

PR

PR 14-MAY-1999; 99US-0134218.

PR

PR 14-MAY-1999; 99US-0134219.

PR

PR 14-MAY-1999; 99US-0134221.

PR

PR 14-MAY-1999; 99US-0134370.

PR

PR 18-MAY-1999; 99US-0134768.

PR

PR 19-MAY-1999; 99US-0134941.

PR

PR 20-MAY-1999; 99US-0135124.

PR

PR 21-MAY-1999; 99US-0135353.

PR

PR 24-MAY-1999; 99US-0135629.

PR

PR 25-MAY-1999; 99US-0136021.

PR

PR 27-MAY-1999; 99US-0136382.

PR

PR 28-MAY-1999; 99US-0136782.

PR

PR 01-JUN-1999; 99US-0137222.

PR

PR 03-JUN-1999; 99US-0137528.

PR

PR 04-JUN-1999; 99US-0137502.

PR

PR 07-JUN-1999; 99US-0137724.

PR

PR 08-JUN-1999; 99US-0138094.

PR

PR 10-JUN-1999; 99US-0138540.  
PR 10-JUN-1999; 99US-0138847.  
PR 14-JUN-1999; 99US-0139119.  
PR 16-JUN-1999; 99US-0139452.  
PR 16-JUN-1999; 99US-0139453.  
PR 17-JUN-1999; 99US-0139492.  
PR 18-JUN-1999; 99US-0139454.  
PR 18-JUN-1999; 99US-0139455.  
PR 18-JUN-1999; 99US-0139456.  
PR 18-JUN-1999; 99US-0139457.  
PR 18-JUN-1999; 99US-0139458.  
PR 18-JUN-1999; 99US-0139459.  
PR 18-JUN-1999; 99US-0139460.  
PR 18-JUN-1999; 99US-0139461.  
PR 18-JUN-1999; 99US-0139462.  
PR 18-JUN-1999; 99US-0139463.  
PR 18-JUN-1999; 99US-0139750.  
PR 18-JUN-1999; 99US-0139763.  
PR 21-JUN-1999; 99US-0139817.  
PR 22-JUN-1999; 99US-0139899.  
PR 23-JUN-1999; 99US-0140353.  
PR 23-JUN-1999; 99US-0140354.  
PR 24-JUN-1999; 99US-0140695.  
PR 28-JUN-1999; 99US-0140823.  
PR 29-JUN-1999; 99US-0140991.  
PR 30-JUN-1999; 99US-0141287.  
PR 01-JUL-1999; 99US-0141842.  
PR 01-JUL-1999; 99US-0142154.  
PR 02-JUL-1999; 99US-0142055.  
PR 06-JUL-1999; 99US-0142390.  
PR 08-JUL-1999; 99US-0142803.  
PR 09-JUL-1999; 99US-0142920.  
PR 12-JUL-1999; 99US-0142977.  
PR 13-JUL-1999; 99US-0143542.  
PR 14-JUL-1999; 99US-0143624.  
PR 15-JUL-1999; 99US-0144005.  
PR 16-JUL-1999; 99US-0144085.  
PR 16-JUL-1999; 99US-0144086.  
PR 19-JUL-1999; 99US-0144325.  
PR 19-JUL-1999; 99US-0144331.  
PR 19-JUL-1999; 99US-0144332.  
PR 19-JUL-1999; 99US-0144333.  
PR 19-JUL-1999; 99US-0144334.  
PR 19-JUL-1999; 99US-0144335.  
PR 20-JUL-1999; 99US-0144352.  
PR 20-JUL-1999; 99US-0144632.  
PR 21-JUL-1999; 99US-0144884.  
PR 21-JUL-1999; 99US-0144814.  
PR 21-JUL-1999; 99US-0145086.  
PR 21-JUL-1999; 99US-0145088.  
PR 22-JUL-1999; 99US-0145085.  
PR 22-JUL-1999; 99US-0145087.  
PR 22-JUL-1999; 99US-0145089.  
PR 22-JUL-1999; 99US-0145192.  
PR 23-JUL-1999; 99US-0145145.  
PR 23-JUL-1999; 99US-0145218.  
PR 23-JUL-1999; 99US-0145224.  
PR 26-JUL-1999; 99US-0145276.  
PR 27-JUL-1999; 99US-0145913.  
PR 27-JUL-1999; 99US-0145918.  
PR 27-JUL-1999; 99US-0145919.  
PR 28-JUL-1999; 99US-0145951.  
PR 02-AUG-1999; 99US-0146386.  
PR 02-AUG-1999; 99US-0146388.  
PR 02-AUG-1999; 99US-0146389.  
PR 03-AUG-1999; 99US-0147038.  
PR 04-AUG-1999; 99US-0147204.  
PR 04-AUG-1999; 99US-0147302.  
PR 05-AUG-1999; 99US-0147192.  
PR 05-AUG-1999; 99US-0147260.  
PR 06-AUG-1999; 99US-0147303.  
PR 06-AUG-1999; 99US-0147416.  
PR 09-AUG-1999; 99US-0147493.

```
PR 09-AUG-1999; 99US-0147935.
PR 10-AUG-1999; 99US-0148171.
PR 11-AUG-1999; 99US-0148319.
PR 12-AUG-1999; 99US-0148341.
PR 13-AUG-1999; 99US-0148565.
PR 13-AUG-1999; 99US-0148684.
PR 16-AUG-1999; 99US-0149368.
PR 17-AUG-1999; 99US-0149175.
PR 18-AUG-1999; 99US-0149426.
PR 20-AUG-1999; 99US-0149722.
PR 20-AUG-1999; 99US-0149723.
PR 20-AUG-1999; 99US-0149929.
PR 23-AUG-1999; 99US-0149902.
PR 23-AUG-1999; 99US-0149930.
PR 25-AUG-1999; 99US-0150566.
PR 26-AUG-1999; 99US-0150884.
PR 27-AUG-1999; 99US-0151065.
PR 27-AUG-1999; 99US-0151066.
PR 27-AUG-1999; 99US-0151080.
PR 30-AUG-1999; 99US-0151303.
PR 31-AUG-1999; 99US-0151438.
PR 01-SEP-1999; 99US-0151930.
PR 07-SEP-1999; 99US-0152363.
PR 10-SEP-1999; 99US-0153070.
PR 13-SEP-1999; 99US-0153758.
PR 15-SEP-1999; 99US-0154018.
PR 16-SEP-1999; 99US-0154039.
PR 20-SEP-1999; 99US-0154779.
PR 22-SEP-1999; 99US-0155139.
PR 23-SEP-1999; 99US-0155486.
PR 24-SEP-1999; 99US-0155659.
PR 28-SEP-1999; 99US-0156458.
PR 29-SEP-1999; 99US-0156596.
PR 04-OCT-1999; 99US-0157117.
PR 05-OCT-1999; 99US-0157753.
PR 06-OCT-1999; 99US-0157865.
PR 07-OCT-1999; 99US-0158029.
PR 08-OCT-1999; 99US-0158232.
PR 12-OCT-1999; 99US-0158369.
PR 13-OCT-1999; 99US-0159293.
PR 13-OCT-1999; 99US-0159294.
PR 13-OCT-1999; 99US-0159295.
PR 14-OCT-1999; 99US-0159329.
PR 14-OCT-1999; 99US-0159330.
PR 14-OCT-1999; 99US-0159331.
PR 14-OCT-1999; 99US-0159637.
PR 14-OCT-1999; 99US-0159638.
PR 18-OCT-1999; 99US-0159584.
PR 21-OCT-1999; 99US-0160741.
PR 21-OCT-1999; 99US-0160767.
PR 21-OCT-1999; 99US-0160768.
PR 21-OCT-1999; 99US-0160770.
PR 21-OCT-1999; 99US-0160814.
PR 21-OCT-1999; 99US-0160815.
PR 22-OCT-1999; 99US-0160980.
PR 22-OCT-1999; 99US-0160981.
PR 22-OCT-1999; 99US-0160989.
PR 25-OCT-1999; 99US-0161404.
PR 25-OCT-1999; 99US-0161405.
PR 25-OCT-1999; 99US-0161406.
PR 26-OCT-1999; 99US-0161359.
PR 26-OCT-1999; 99US-0161360.
PR 26-OCT-1999; 99US-0161361.
PR 28-OCT-1999; 99US-0161920.
PR 28-OCT-1999; 99US-0161992.
PR 28-OCT-1999; 99US-0161993.
PR 29-OCT-1999; 99US-0162142.

Query Match 36.0%; Score 123; DB 21; Length 624;
Best Local Similarity 68.1%; Pred. No. 7.5e-32;
Matches 171; Conservative 0; Mismatches 80; Indels 0; Gaps 0;
QY 23 gccacaatgattgagtggtttgcaacgccgtcttaggaagaagtcgcgcgttaagtgc 82

Db 127 gcgaagatgacgaggtggtgctgaacgaccgctcggaagaagtgctggtggaagtcgc 186
QY 83 aacaccgatgacaccatcggtgacttgaagaactgatagcgcccaactggcactcgt 142
Db 187 aacgaggacgacacatcgcgacctcaagaagctggcgcgcgacagcgagcgacgc 246
QY 143 tggataagatcgcttcttaaaagtgtacacgatttttaaggagaccatgtatctctgga 202
Db 247 ccgagaagatccgcacatccagaagtggtacacatctacaaggagaccatcacgtccaag 306
QY 203 gattatgaatccacgatgggatgaacctggagctttattaccactagagggaattctc 262
Db 307 gactacgagatccacgacgggatgggtcgctgagctctactacaactgagcgccctgaccc 366
QY 263 ccaccttgccc 273
Db 367 caaccgtgcc 377

RESULT 10
C47007
ID C47007 standard; DNA; 556 BP.
AC C47007;
XX
DT 18-OCT-2000 (first entry)
XX
DE Arabidopsis thaliana DNA fragment SEQ ID NO: 52220.
XX
XX Hybridisation assay; genetic mapping; gene expression control;
KW protein identification; signal transduction pathway;
KW metabolic pathway; promoter; termination sequence; ss.
XX
XX Arabidopsis thaliana.
PN EP1033405-A2.
XX
PD 06-SEP-2000.
XX
PF 25-FEB-2000; 2000EP-0301439.
XX
PR 25-FEB-1999; 99US-0121825.
PR 05-MAR-1999; 99US-0123180.
PR 09-MAR-1999; 99US-0123548.
PR 23-MAR-1999; 99US-0125788.
PR 25-MAR-1999; 99US-0126284.
PR 29-MAR-1999; 99US-0126785.
PR 01-APR-1999; 99US-0127462.
PR 06-APR-1999; 99US-0128234.
PR 08-APR-1999; 99US-0128714.
PR 16-APR-1999; 99US-0129845.
PR 19-APR-1999; 99US-0130077.
PR 21-APR-1999; 99US-0130449.
PR 23-APR-1999; 99US-0130510.
PR 28-APR-1999; 99US-0130891.
PR 30-APR-1999; 99US-0131449.
PR 30-APR-1999; 99US-0132048.
PR 30-APR-1999; 99US-0132407.
PR 04-MAY-1999; 99US-0132484.
PR 05-MAY-1999; 99US-0132485.
PR 06-MAY-1999; 99US-0132486.
PR 06-MAY-1999; 99US-0132487.
PR 07-MAY-1999; 99US-0132863.
PR 11-MAY-1999; 99US-0134256.
PR 14-MAY-1999; 99US-0134218.
PR 14-MAY-1999; 99US-0134219.
PR 14-MAY-1999; 99US-0134221.
PR 14-MAY-1999; 99US-0134370.
PR 18-MAY-1999; 99US-0134768.
PR 19-MAY-1999; 99US-0134941.
PR 20-MAY-1999; 99US-0135124.
PR 21-MAY-1999; 99US-0135353.
```





```

PR 27-JUL-1999; 99US-0145918.
PR 27-JUL-1999; 99US-0145919.
PR 28-JUL-1999; 99US-0145951.
PR 02-AUG-1999; 99US-0146386.
PR 02-AUG-1999; 99US-0146388.
PR 02-AUG-1999; 99US-0146389.
PR 03-AUG-1999; 99US-0147038.
PR 04-AUG-1999; 99US-0147204.
PR 04-AUG-1999; 99US-0147302.
PR 05-AUG-1999; 99US-0147192.
PR 05-AUG-1999; 99US-0147260.
PR 06-AUG-1999; 99US-0147303.
PR 06-AUG-1999; 99US-0147416.
PR 09-AUG-1999; 99US-0147493.
PR 09-AUG-1999; 99US-0147935.
PR 10-AUG-1999; 99US-0148171.
PR 11-AUG-1999; 99US-0148319.
PR 12-AUG-1999; 99US-0148341.
PR 13-AUG-1999; 99US-0148565.
PR 13-AUG-1999; 99US-0148684.
PR 16-AUG-1999; 99US-0149368.
PR 17-AUG-1999; 99US-0149175.
PR 18-AUG-1999; 99US-0149426.
PR 20-AUG-1999; 99US-0149722.
PR 20-AUG-1999; 99US-0149723.
PR 20-AUG-1999; 99US-0149929.
PR 23-AUG-1999; 99US-0149902.
PR 23-AUG-1999; 99US-0149930.
PR 25-AUG-1999; 99US-0150566.
PR 26-AUG-1999; 99US-0150894.
PR 27-AUG-1999; 99US-0151065.
PR 27-AUG-1999; 99US-0151066.
PR 27-AUG-1999; 99US-0151080.
PR 30-AUG-1999; 99US-0151303.
PR 31-AUG-1999; 99US-0151438.
PR 01-SEP-1999; 99US-0151930.
PR 07-SEP-1999; 99US-0152363.
PR 10-SEP-1999; 99US-0153070.
PR 13-SEP-1999; 99US-0153758.
PR 15-SEP-1999; 99US-0154018.
PR 16-SEP-1999; 99US-0154039.
PR 20-SEP-1999; 99US-0154779.
PR 22-SEP-1999; 99US-0155139.
PR 23-SEP-1999; 99US-0155486.
PR 24-SEP-1999; 99US-0155659.
PR 28-SEP-1999; 99US-0156458.
PR 29-SEP-1999; 99US-0156596.
PR 04-OCT-1999; 99US-0157117.
PR 05-OCT-1999; 99US-0157753.
PR 06-OCT-1999; 99US-0157865.
PR 07-OCT-1999; 99US-0158029.
PR 08-OCT-1999; 99US-0158232.
PR 12-OCT-1999; 99US-0158369.
PR 13-OCT-1999; 99US-0159293.
PR 13-OCT-1999; 99US-0159294.
PR 13-OCT-1999; 99US-0159295.
PR 14-OCT-1999; 99US-0159329.
PR 14-OCT-1999; 99US-0159330.
PR 14-OCT-1999; 99US-0159331.
PR 14-OCT-1999; 99US-0159637.
PR 14-OCT-1999; 99US-0159638.
PR 18-OCT-1999; 99US-0159584.
PR 21-OCT-1999; 99US-0160741.
PR 21-OCT-1999; 99US-0160767.
PR 21-OCT-1999; 99US-0160768.
PR 21-OCT-1999; 99US-0160770.
PR 21-OCT-1999; 99US-0160814.
PR 21-OCT-1999; 99US-0160815.
PR 22-OCT-1999; 99US-0160980.
PR 22-OCT-1999; 99US-0160981.
PR 22-OCT-1999; 99US-0160989.
PR 25-OCT-1999; 99US-0161404.
PR 25-OCT-1999; 99US-0161405.

PR 25-OCT-1999; 99US-0161406.
PR 26-OCT-1999; 99US-0161359.
PR 26-OCT-1999; 99US-0161360.
PR 26-OCT-1999; 99US-0161361.
PR 28-OCT-1999; 99US-0161920.
PR 28-OCT-1999; 99US-0161992.
PR 28-OCT-1999; 99US-0161993.
PR 29-OCT-1999; 99US-0162142.

Query Match 32.5%; Score 111.2; DB 21; Length 472;
Best Local Similarity 69.1%; Pred. No. 6.6e-28;
Matches 152; Conservative 0; Mismatches 68; Indels 0; Gaps 0;

QY 29 atgattgaggtgttgcaacgaccgtctagaaagaagtcgcttaagtgaacacc 88
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 148 atgacgaggtgtgctcaacgaccgtctggaagaagtgctggaagtcgaacgag 207
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 89 gatgacaccatcggtgactgaagaactgacgagcccaactgacactgcttgaat 148
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 208 gacgacaccatcggtgactgaagaactgacgagcccaactgacactgcttgaat 267
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 149 aagatcgttcttaaaagtgtacacgattttaagaccatgtatctctggagattat 208
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 268 aagatccgcacatccagaagtggtacaacatctacaaggactacatcaccctcaaggactac 327
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 209 gaaatccacgatgggatgaacctggagctttattaccagt 248
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 328 gaggtccacgacggcatggcctcgactctactacaact 367
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

RESULT 12
A31662
ID A31662 standard; DNA; 460 BP.
XX
AC A31662;
XX
DT 05-JUL-2000 (first entry)
XX
DE Plant microsatellite marker #623.
XX
KW Plant microsatellite sequence; core repeat sequence; detection; probe;
KW DNA polymorphism; genome mapping; physical mapping; fingerprinting;
KW variety identification; genetic variability evaluation; primer; ss.
XX
OS Eucalyptus grandis.
XX
PN WO9967421-A1.
XX
PD 29-DEC-1999.
XX
PF 25-JUN-1999; 99WO-NZ00092.
XX
PR 25-JUN-1998; 98US-0105307.
XX
PA (GENE-) GENESIS RES & DEV CORP LTD & FLETCHER.
PA (FLET-) FLETCHER CHALLENGE FORESTS LTD.
XX
PI Havukkala IJ, Bloksberg LN, Glenn M;
XX
DR WPI; 2000-116958/10.
XX
PT New plant microsatellite markers and associated flanking species for
PT the detection of polymorphic genetic markers -
XX
PS Claim 1; Page 260; 392pp; English.
XX
CC Sequences A31040-A32093 represent novel plant microsatellite sequences
CC and associated flanking species. The sequences comprise a central core
CC repeat sequence, especially selected from the sequences A32094-A32096
CC with left and right flanking sequences. The polynucleotide sequences
CC can be used in the detection of DNA polymorphisms, in genome mapping,
CC in physical mapping, in positional cloning of genes, in variety
CC identification and in evaluation of genetic variability within and

```

CC between plant tissues, populations, cultivars, species and species  
 CC groups. They may also be used to design hybridization probes for  
 CC oligonucleotide fingerprinting and library screening and to design  
 CC primers for microsatellite-primed PCR. Microsatellite markers are  
 CC useful to locate specific economically useful genes in plant genomes.  
 XX  
 SQ Sequence 460 BP; 122 A; 116 C; 146 G; 76 T; 0 other;

Query Match 30.6%; Score 104.6; DB 21; Length 460;  
 Best Local Similarity 69.1%; Pred. No. 1.1e-25;  
 Matches 143; Conservative 0; Mismatches 64; Indels 0; Gaps 0;  
 QY 20 caagccacaatgattgaggtgttgcacacccgtctaggaagaagtcgcgttaag 79  
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||  
 Db 227 caagcgaatgatagaggtgtgctgacacacccgctggggaaggtgaggtggaag 286  
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||  
 QY 80 tgaacacccgatgacacacccgtgggagcttgaaagaactgatacgcccaactggcact 139  
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||  
 Db 287 tgaacacccgatgacacacccgtgggagcttgaaagaactgatacgcccaactggcact 346  
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||  
 QY 140 cgttgaataagatcgttcttcttaaaagtgtacacagatttttaaggaccatgtatctctg 199  
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||  
 Db 347 cgcgcgacaaagatccgaatccagaagtgtacacatctctacaaaggaccacacccctc 406  
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||  
 QY 200 ggagattatgaatccacgatggatg 226  
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||  
 Db 407 aaggactacgaggtccacgacggcatg 433

RESULT 13  
 X57370  
 ID X57370 standard; DNA; 102 BP.  
 AC  
 AC X57370;  
 DT 24-JUL-1999 (first entry)  
 XX Human beacon DNA.  
 XX  
 KW Beacon; hypothalamus; obese; lean; agonist; antagonist; treatment;  
 KW obesity; anorexia; weight maintenance; energy imbalance; diabetes;  
 KW metabolic syndrome; dyslipidemia; hypertension; insulin resistance;  
 KW medicament; livestock; diagnosis; human; ss.  
 XX  
 OS Homo sapiens.  
 OS WO9223217-A1.  
 XX  
 PD 14-MAY-1999.  
 XX  
 PF 30-OCT-1998; 98WO-AU00902.  
 XX  
 PR 11-NOV-1997; 97AU-0000323.  
 PR 31-OCT-1997; 97AU-0000117.  
 XX  
 PA (ITDI-) INT DIABETES INST.  
 PA (UYDE-) UNIV DEAKIN.  
 XX  
 PI Collier G, Zimmet P;  
 XX  
 DR WPI; 1999-337484/28.  
 DR P-PSDB; Y08414.  
 XX  
 PT New gene encoding a beacon protein associated with modulation of  
 PT obesity, diabetes and metabolic energy levels  
 XX  
 PS Claim 3; Page 53; 85pp; English.  
 XX  
 CC This invention describes a novel beacon protein and its encoding nucleic  
 CC acid which is expressed in larger amounts in hypothalamus tissue of obese  
 CC animals compared to lean animals. Agonists and antagonists of beacon can  
 CC be used to treat obesity, anorexia, weight maintenance, energy imbalance,

CC diabetes, metabolic syndrome, dyslipidemia, hypertension and/or insulin  
 CC resistance. The beacon protein, itself is used to manufacture medicaments  
 CC for treatment of obesity, anorexia, energy imbalance or diabetes. The  
 CC treatment is contemplated for both human and animals, such as those  
 CC important to the livestock industry. The antibody and polynucleotides are  
 CC useful in diagnosis of conditions as above.  
 XX  
 SQ Sequence 102 BP; 30 A; 21 C; 28 G; 22 T; 1 other;

Query Match 22.5%; Score 77; DB 20; Length 102;  
 Best Local Similarity 84.3%; Pred. No. 1.3e-16;  
 Matches 86; Conservative 0; Mismatches 16; Indels 0; Gaps 0;  
 QY 29 atgattgaggtgttgcacacccgtctaggaagaagtcgcgttaagtgcacacc 88  
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||  
 Db 1 atgacgaggtgttgcacacccgtctgggaaaggctccncttaaatgcaacacg 60  
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||  
 QY 89 gatgacaccatcggtgacttgaaagaactgatacgcgcccaa 130  
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||  
 Db 61 gatgataccatcggtgacttgaaagaactgatacgagcctaa 102  
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

RESULT 14  
 C81770  
 ID C81770 standard; DNA; 102 BP.  
 XX  
 AC C81770;  
 XX  
 DT 23-FEB-2001 (first entry)  
 XX Human beacon (short form) coding sequence.  
 XX  
 DE Israel sand rat; beacon; obesity; NIDDM; energy balance; diabetes;  
 KW ligand; human; ds.  
 KW  
 XX Homo sapiens.  
 XX WO200064931-A1.  
 XX  
 PD 02-NOV-2000.  
 XX  
 PF 19-APR-2000; 2000WO-AU00342.  
 XX  
 PR 23-APR-1999; 99AU-0009919.  
 PR 24-MAR-2000; 2000AU-0006454.  
 XX  
 PA (AUTO-) AUTOGEN PTY LTD.  
 XX  
 PI Collier G, Walder K, Zimmet P;  
 XX  
 DR WPI; 2000-687311/67.  
 DR P-PSDB; B36291.  
 XX  
 PT Ligand of beacon protein useful for treating obesity, anorexia, energy  
 PT imbalance, diabetes, metabolic syndrome, dyslipidemia, hypertension and  
 PT insulin resistance  
 XX  
 PS Claim 3; Fig 1B; 67pp; English.  
 XX  
 CC The present invention is related to the isolation of a ligand known as  
 CC beacon from the Israeli sand rat. Beacon is associated with the  
 CC regulation of energy balance, and the protein, its coding sequence and  
 CC analogues can be used in the treatment of diabetes, obesity, anorexia,  
 CC energy imbalance, metabolic syndrome, dyslipidemia, hypertension and  
 CC insulin resistance. In addition, they can be used in agriculture to  
 CC produce leaner animals.  
 CC Note: This sequence is stated in the claims as being a protein sequence,  
 CC rather than a nucleic acid.  
 XX  
 SQ Sequence 102 BP; 30 A; 21 C; 28 G; 22 T; 1 other;

